PAPERS AND PROCEEDINGS

OF THE

Seventy-ninth Annual Meeting

OF THE

AMERICAN ECONOMIC ASSOCIATION

San Francisco, California, December 27-29, 1966

Edited by the Secretary of the Association

AMERICAN ECONOMIC ASSOCIATION

Organized at Saratoga, New York, September 9, 1885

PAST OFFICERS

Presidents

1. FRANCIS A. WALKER,* M.I.T., 1886-92
2. CHARLES F. DUNBAR,* Harvard, 1893
3. JOHN B. CLARK,* Columbia, 1894-95
4. HENRY C. ADAMS,* Michigan, 1896-97
5. ARTHUR T. HADLEY,* Yale, 1898-99
6. RICHARD T. ELY,* Wisconsin, 1900-01
7. EDWIN R. A. SELIGMAN,* Columbia, 1902-03
8. FRANK W. TAUSSIG,* Harvard, 1904-05
9. JEREMIAH W. JENKS,* Cornell, 1906-07
10. SIMON N. PATTEN,* Pennsylvania, 1908
11. DAVIS R. DEWEY,* M.I.T., 1909
12. EDMUND J. JAMES,* Illinois, 1910
13. HENRY W. FARNAM,* Yale, 1911
14. FRANK A. FETTER,* Princeton, 1912
15. DAVID KINLEY,* Illinois, 1913
16. JOHN H. GRAY,* Minnesota, 1914
17. WALTER F. WILLCOX,* Cornell, 1915
18. THOMAS N. CARVER,* Harvard, 1916
19. JOHN R. COMMONS,* Wisconsin, 1917
20. IRVING FISHER,* Yale, 1918
21. HENRY B. GARDNER,* Brown, 1919
22. HERBERT J. DAVENPORT,* Cornell, 1920
23. JACOB H. HOLLANDER,* Johns Hopkins, 1921
24. HENRY R. SEAGER,* Columbia, 1922
25. CARL C. PLEHN,* California, 1923
26. WESLEY C. MITCHELL,* Columbia, 1924
27. ALLYN A. YOUNG,* Harvard, 1925
28. EDWIN W. KEMMERER,* Princeton, 1926
29. THOMAS S. ADAMS,* Yale, 1927
30. FRED M. TAYLOR,* Michigan, 1928
31. EDWIN F. GAY,* Harvard, 1929
32. MATTHEW B. HAMMOND,* Ohio State, 1930
33. ERNEST L. BOGART,* Illinois, 1931
34. GEORGE E. BARNETT,* Johns Hopkins, 1932
35. WILLIAM Z. RIPLEY,* Harvard, 1933
36. HARRY A. MILLIS,* Chicago, 1934
37. JOHN M. CLARK,* Columbia, 1935
38. ALVIN S. JOHNSON, New School, 1936
39. OLIVER M. W. SPRAGUE,* Harvard, 1937
40. ALVIN H. HANSEN, Harvard, 1938
41. JACOB VINER, Chicago, 1939
42. FREDERICK C. MILLS,* Columbia, 1940
43. SUMNER H. SLICHTER,* Harvard, 1941
44. EDWIN G. NOURSE, Brookings, 1942
45. ALBERT B. WOLFE, Ohio State, 1943

46. JOSEPH S. DAVIS, Stanford, 1944
47. I. L. SHARFMAN,* Michigan, 1945
48. E. A. GOLDENWEISER,* Institute for Advanced Study, 1946
49. PAUL H. DOUGLAS, Chicago, 1947
50. JOSEPH A. SCHUMPETER,* Harvard, 1948
51. HOWARD S. ELLIS, California, 1949
52. FRANK H. KNIGHT, Chicago, 1950
53. JOHN H. WILLIAMS, Harvard, 1951
54. HAROLD A. INNIS,* Toronto, 1952
55. CALVIN B. HOOVER, Duke, 1953
56. SIMON KUZNETS, Pennsylvania, 1954
57. JOHN D. BLACK,* Harvard, 1955
58. EDWIN E. WITTE,* Wisconsin, 1956
59. MORRIS A. COPELAND, Cornell, 1957
60. GEORGE W. STOCKING, Vanderbilt, 1958
61. ARTHUR F. BURNS, Columbia, 1959
62. THEODORE W. SCHULTZ, Chicago, 1960
63. PAUL A. SAMUELSON, M.I.T., 1961
64. EDWARD S. MASON, Harvard, 1962
65. GOTTFRIED HABERLER, Harvard, 1963
66. GEORGE J. STIGLER, Chicago, 1964
67. JOSEPH J. SPENGLER, Duke, 1965
68. FRITZ MACHLUP, Princeton, 1966

Secretaries

RICHARD T. ELY,* 1886-92
EDWARD A. ROSS,* 1893
JEREMIAH W. JENKS,* 1894-96
WALTER F. WILLCOX,* 1897-99

Treasurers

EDWIN R. A. SELIGMAN,* 1886-90
FREDERICK B. HAWLEY,* 1891-95
CHARLES H. HULL,* 1896-99

Secretary-Treasurers

CHARLES H. HULL,* 1900
FRANK A. FETTER,* 1901-06
WINTHROP M. DANIELS,* 1907-08
THOMAS N. CARVER,* 1909-13
ALLYN A. YOUNG,* 1914-20
RAY B. WESTERFIELD,* 1921-25
FREDERICK S. DEIBLER,* 1925-35
JAMES WASHINGTON BELL,* 1936-61

Editors

DAVIS R. DEWEY,* 1911-40
PAUL T. HOMAN, 1941-51
BERNARD F. HALEY, 1952-62
* Deceased.

OFFICERS FOR 1967

President
MILTON FRIEDMAN, Chicago

President-elect
KENNETH E. BOULDING, Michigan

Vice-Presidents
WILLIAM J. BAUMOL, Princeton
ROBERT TRIFFIN, Yale

Secretary-Treasurer and Editor of PROCEEDINGS
HAROLD F. WILLIAMSON, Northwestern

Managing Editor of the AMERICAN ECONOMIC REVIEW
JOHN G. GURLEY, Stanford

Managing Editor of the JOURNAL OF ECONOMIC ABSTRACTS
ARTHUR SMITHIES, Harvard

Elected Members of the Executive Committee
MOSES ABRAMOVITZ, Stanford
NEIL H. JACOBY, California, Los Angeles
JAMES M. BUCHANAN, Virginia
LAWRENCE R. KLEIN, Pennsylvania
KENNETH J. ARROW, Stanford
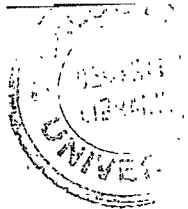FRANCO MODIGLIANI, M.I.T.

Ex Officio Members
JOSEPH J. SPENGLER, Duke
FRITZ MACHLUP, Princeton

Counsel
SEYMOUR J. RUBIN, Washington, D.C.

PAPERS AND PROCEEDING

OF THE

Seventy-ninth Annual Meeting

OF THE

AMERICAN ECONOMIC ASSOCIATION

San Francisco, California, December 27-29, 1966

TABLE OF CONTENTS

PROCEEDINGS

PROGRAM OF THE SEVENTY-NINTH ANNUAL MEETING OF THE
AMERICAN ECONOMIC ASSOCIATION

San Francisco, California, December 27-29, 1966

Monday, December 26, 1966

3:00 P.M.

Executive Committee Meeting

Tuesday, December 27, 1966

8:30 A.M.

Economic Analysis of Water Resource Problems (Joint session with the American Farm
Economic Association)
  Chairman: MAURICE M. KELSO, University of Arizona
  Papers: A. ALLAN SCHMID, Michigan State University; JACK HIRSHLEIFER, University
  of California, Los Angeles, and J. W. MILLIMAN, Indiana University; S. V. CIRIACY-
  WANTRUP, University of California, Berkeley
  Discussion: ROBERT H. HAVEMAN, Grinnell College; MASON GAFFNEY, University of
  Wisconsin; ARTHUR MAASS, Harvard University

The Measurement of Price Change
  Chairman: DANIEL BRILL, Board of Governors of the Federal Reserve System
  Papers: IRVING B. KRAVIS, University of Pennsylvania, and ROBERT E. LIPSEY, National
  Bureau of Economic Research; JAMES K. KINDAHL, University of Chicago; PHOEBUS
  J. DHRYMES, University of Pennsylvania
  Discussion: DOROTHY S. BRADY,[1] University of Pennsylvania; GREGORY KIPNIS, National
  Industrial Conference Board
Conditions for Technological Progress and Economic Growth in Communist China
  (Joint session with the Association for the Study of Soviet-type Economies)
  Chairman: ALEXANDER ECKSTEIN, University of Michigan
  Papers: EDWIN F. JONES, Department of State; J. C. LIU, McGill University; ROBERT
  DERNBERGER, University of Chicago
  Discussion: FRANKLIN D. HOLZMAN, Tufts University; HENRY ROSOVSKY, Harvard
  University
Graduate Student Papers (Joint session with *The American Economist*)
  Chairman: GERHARD TINTNER, University of Southern California
  Papers: RASESH THAKKAR, University of Rochester; ELEANOR M. BIRCH and JOHN M.
  HEINEKE, University of Iowa; ALEX G. VICAS, Loyola College (Canada)
  Discussion: MARK Z. FABRYCY, New York University; JOHN BOORMAN; University of
  Southern California; ERNEST J. BARTELL, Notre Dame University

10:45 A.M.

Policy Round Table 1: Some Problems of Full Employment (Joint session with the
S.S.R.C. Committee on Economic Stability)
  Moderator: R. A. GORDON, University of California, Berkeley
  Panel: WALTER HOADLEY, Bank of America; JOSEPH PECHMAN, Brookings Institution;
  WARREN L. SMITH, University of Michigan; HERBERT STEIN, Committee for Economic
  Development
Policy Round Table 2: Research and Development
  Moderator: H. E. RILEY, National Science Foundation
  Panel: BURTON KLEIN, RAND Corporation; ZVI GRILICHES, University of Chicago

2:30 P.M.

Military Manpower Procurement (Joint session with Industrial Relations Research Associ-
ation)
  Chairman: HAROLD WOOL, U.S. Department of Defense
  Papers: STUART H. ALTMAN, Brown University, and ALAN E. FECHTER, Institute of De-

[1] No manuscript received.

fense Analyses; FRANKLIN M. FISHER, Massachusetts Institute of Technology, and ANTON S. MORTON, Arthur D. Little, Inc.; WALTER Y. OI, University of Washington
Discussion: GEORGE H. HILDEBRAND, Cornell University; PAUL A. WEINSTEIN, University of Maryland; HAROLD WOOL, U.S. Department of Defense
Economic Development (Joint session with Econometric Society)
Chairman: MOSES ABRAMOVITZ, Stanford University
Papers: RONDO E. CAMERON, University of Wisconsin; EDWARD F. DENISON, Brookings Institution; HOLLIS B. CHENERY and ROBERT DORFMAN,[1] Harvard University
Discussion: HENRY ROSOVSKY, Harvard University; CHARLES P. KINDLEBERGER, Massachusetts Institute of Technology; EMILE DESPRES,[1] Stanford University
Aggregate Models
Chairman: SAMUEL H. NERLOVE, University of California, Los Angeles
Papers: CARL F. CHRIST, Johns Hopkins University; HAROLD SHAPIRO, University of Michigan; EDGAR L. FEIGE, University of Wisconsin
Discussion: GARY FROMM, Brookings Institution; THOMAS R. SAVING, Michigan State University; GEORGE R. MORRISON, State University of New York, Buffalo

8:00 P.M.
Richard T. Ely Lecture
Chairman: THEODORE W. SCHULTZ, University of Chicago
Invited Lecture: ABBA P. LERNER, University of California, Berkeley

Wedneday, December 28, 1966
8:30 A.M.
Cost-Benefit Analysis for Government Decisions
Chairman: WILLIAM H. MECKLING, University of Rochester
Papers: STEPHEN ENKE; WORTH BATEMAN, U.S. Department of Health, Education and Welfare; WILLIAM ROSS, U.S. Department of Housing and Urban Development
Discussion: GEORGE W. HILTON, University of California, Los Angeles; ARNOLD MOORE, U.S. Department of Defense; MARTIN ANDERSON, Columbia University; HENRY ROWEN, Bureau of the Budget
Property Rights and Behavior
Chairman: ARMEN A. ALCHIAN, University of California, Los Angeles
Papers: ALFRED NICOLS, University of California, Los Angeles; HAROLD DEMSETZ, University of Chicago; OTTO DAVIS, Carnegie Institute of Technology, and ANDREW WHINSTON, Purdue University
Discussion: JOHN ASHLEY, California State College, Hayward; PAUL L. KLEINSORGE, University of Oregon; HOWARD KUNREUTHER, University of Chicago
Economic History
Chairman: ARTHUR H. COLE, Harvard University
Papers: EARL J. HAMILTON, University of Chicago and State University of New York, Binghamton; WARREN J. SCOVILLE, University of California, Los Angeles; PAUL DAVID, Stanford University
Discussion: JOHN H. DALE,[1] University of Toronto; STANLEY ENGERMAN, University of Rochester; ROSS M. ROBERTSON, Indiana University

10:45 A.M.
Policy Round Table 3: Guideposts and Inflation (Joint session with Industrial Relations Research Association)
Moderator: ALBERT REES, Princeton University
Panel: WILLIAM G. BOWEN, Princeton University; JOHN T. DUNLOP, Harvard University; OTTO ECKSTEIN, Harvard University; GEORGE SCHULTZ, University of Chicago
Policy Round Table 4: Regulation of Industry
Moderator: NEIL H. JACOBY, University of California, Los Angeles
Panel: WILLIAM J. BAUMOL, Princeton University; JOEL DEAN, Columbia University; JOHN R. MEYER, Harvard University; GEORGE J. STIGLER, University of Chicago

12:30 P.M.
Joint Luncheon with American Finance Association
Chairman: FRITZ MACHLUP, Princeton University
Speaker: ALLAN SPROUL, California

2:30 P.M.

Antitrust and Monopoly

    Chairman: AARON DIRECTOR, University of Chicago
    Papers: ROBERT H. BORK, Yale University; WILLIAM COMANOR, Harvard University;
       WILLIAM HAMBURGER, University of North Carolina
    Discussion: PHIL C. NEAL,[1] University of Chicago; JOHN S. McGEE, Duke University;
       B. PETER PASHIGIAN, University of Chicago

Invited Doctoral Dissertations I

    Chairman: E. CAREY BROWN, Massachusetts Institute of Technology
    Papers: JAMES N. ROSSE, Stanford University; MIGUEL SIDRAUSKI, Massachusetts Insti-
       tute of Technology; JOHN PIPPENGER, University of California, Santa Barbara
    Discussion: HARRY EASTMAN, University of Toronto; DALE W. JORGENSON, University
       of California, Berkeley; PETER O. STEINER, University of Wisconsin

Topics in Money (Joint session with American Finance Association)

    Chairman: J. FRED WESTON, University of California, Los Angeles
    Papers: DAVID FAND, State University of New York, Buffalo; A. LEIJONHUFVUD, Univer-
       sity of California, Los Angeles; WILLIAM BRAINARD, Yale University
    Discussion: ALLAN H. MELTZER, Carnegie Institute of Technology; MYRON H. ROSS,
       Western Michigan University; E. J. KANE, Boston College

8:00 P.M.

Presidential Address[2]

    Chairman: RICHARD LESTER, Princeton University
    Presidential Address: FRITZ MACHLUP, Princeton University

9:15 P.M.

Business Meeting

Thursday, December 29, 1966

8:30 A.M.

Economics of Health

    Chairman: JOSEPH J. SPENGLER, Duke University
    Papers: SIMON ROTTENBERG, Duke University; MILLARD F. LONG, University of Chicago,
       and PAUL J. FELDSTEIN, University of Michigan; ROBIN BARLOW, University of
       Michigan
    Discussion: GEORGE H. BORTS, Brown University; HERBERT E. KLARMAN, Johns Hopkins
       University; PETER NEWMAN, Johns Hopkins University

*The "Convergence" Hypothesis: The Planning and Market Elements in the Develop-
    ment of Soviet, Western, and Developing Economies* (Joint session with Associa-
    tion for Comparative Economics)

    Chairman: GERHARD COLM, National Planning Association
    Papers: ABRAM BERGSON, Harvard University; NEIL CHAMBERLIN, Yale University;
       SHANTI S. TANGRI, Wayne State University
    Discussion: VICTOR PERLO, Economic Consultant; ALLAN GRUCHY, University of Mary-
       land; ARTHUR LERMER, Sir George Williams University

Invited Doctoral Dissertations II

    Chairman: JOE S. BAIN, University of California, Berkeley
    Papers: SEIJI NAYA, University of Wisconsin; WILLIAM F. CAMPBELL, Louisiana State
       University; WILLIAM LANDES, University of Chicago
    Discussion: O. H. BROWNLEE,[1] University of Minnesota; RICHARD NELSON, RAND Cor-
       poration; ANTHONY M. TANG, Vanderbilt University and University of Hong Kong

10:45 A.M.

Policy Round Table 5: International Payments Problems

    Moderator: GOTTFRIED HABERLER, Harvard University
    Panel: EMILE DESPRES, Stanford University; H. G. JOHNSON, University of Chicago and
       London School of Economics; PETER KENEN, Columbia University; ROBERT ROOSA,
       Brown Brothers Harriman and Company

---

[2] Published in the March, 1967, *A.E.R.*

RICHARD T. ELY LECTURE

EMPLOYMENT THEORY AND EMPLOYMENT POLICY

By ABBA P. LERNER
University of California, Berkeley

In the early years of the Keynesian era, before all the anti-Keynesians became Keynesian anti-Keynesians, we knew that we could maintain full employment while avoiding inflation by having a rate of effective demand that was neither too small nor too large. All that remained was to get the authorities to do the right thing. Those who spoke of inflationary dangers either were prisoners of the inappropriate assumption of full employment or saw in the expansion of demand an infernal "sorcerer's apprentice" machine with no "stop" button.

Now, in the fourth decade of the Keynesian era, we have graduated from a state of innocence of the inflationary dangers from full employment policies to a state of obsession with them.

This is not due to a lack of theories of inflation. We even have models with price levels rising or falling in any degree but identical in real terms, the different price changes being offset by different money rates of interest and adjustments in monetary obligations. But we generally make price stability a basic objective, so that degrees of inflation are automatically degrees of evil. We have blamed the evil on too much money, on too much spending, on increases in these, on such increases relative to the volume of goods to be bought, and on an excess of investment over saving. But all of these have about the same intent, for which I shall use the term "excess demand." This has the additional convenience that it can also be blamed for "repressed inflation" when prices do not rise.

We have also blamed excess demand on expectations of rising prices, and blamed these expectations on expectations of increases in incomes and so on, but our problem today is not in the search for causes of causes. It is the breakdown of the simple Keynesian formula for preventing both inflation and unemployment by avoiding both excess demand and deficient demand.

This breakdown comes from an underestimation of the Keynesian revolution. The key element in the revolution was the official recognition of a downward rigidity of the money wage. Its significance was obscured at first by Keynes's focusing on the inability of workers to reduce the real wage. The confusion was compounded by Keynes's

1

denial that downward wage flexibility would cure the unemployment and a great debate between neoclassical economists arguing that with sufficient wage flexibility falling prices would cure any unemployment by increasing the real value of claims to wealth, and Keynesians pointing out that Keynes spoke only of degrees of flexibility achievable in the real world, which would be swamped by deflationary dynamic expectations. Both sides were right and both failed to see that the recognition of "downward rigidity" involved nothing less than the dethronement of the market as the determinant of price!

If supply exceeds demand, the market is telling the price to fall. If the price does not fall, it is obeying a more powerful master—a price administrator who can just as well tell the wage to go up when the market is telling it to go down. And this is exactly what we see in the un-Keynesian coexistence of unemployment with inflation—un-Keynesian because Keynes not only did not see this possibility himself but even succeeded in banishing it from the vision of others.

In 1935 or 1936, Robert Bryce and I asked Keynes whether the maintenance of sufficient effective demand for full employment could not result in perpetuating inflation. Keynes could not understand our question, although we persisted until we exhausted his patience—and apparently also my own, for the question of administered inflation seems to have been washed from my brain for a decade or two by this episode.

Cost inflation or administered inflation or markup inflation is now receiving much attention. As 'wages rise, prices rise too by virtue of the markup of price over cost by the price administrators of business. The wage administrators then raise the wage again to restore its real value or its markup over the minimum basic cost of living. The two sets of markup administrators, trying to get round each other, provide the twist or couple that drives the inflation.

Just like demand inflation, administered inflation persists by virtue of ever renewed attempts at the impossible. In demand inflation the attempt is to buy more than 100 percent of the supply. In administered inflation it is the attempt of labor and business, by their respective markups, to get, between them, more than 100 percent of the joint product.

Special credit should go to Professor Sidney Weintraub for his attempts to direct the attention of economists to this crucial markup of price over wage or ratio of the national product to the wage bill, instead of debating the relative inadequacies of the liquidity-preference half and the time-preference half of the theory of effective demand—a debate not indeed on the famous question as to which blade of the scissors does the cutting, but which blade would make a worse razor.

Noting that the price-wage ratio was more stable than either the velocity of circulation or the propensity to consume, Weintraub unfortunately could not resist the temptation to provide it with its own truistic "basic formula" to rival the quantity equation and the multiplier that are in the service of the other ratios.

There are, however, those who are so impressed by the efficiency of the whole scissors, even though neither blade makes a good razor, that they are unable to put it down even when a different instrument is called for. They try to cure administered inflation by reducing effective demand because that works for demand inflation, and then they try to cure the resulting unemployment by increasing effective demand, back and forth, in what the British have dubbed "Stop and Go." Disappointed, frustrated, and disgusted with the inevitable failure, they declare that Keynes is all wet and functional finance a mirage.

In my first encounter with the indecent entrance of the inflation dragon on the stage before his cue was called, i.e., before full employment was reached, some remnant of the brainwashing made me give the name "full employment" (albeit "low full employment") to the point where the dragon first appeared—whispering the cue, as it were, to legitimize the dragon's prematurity. But I nevertheless still find this a convenient expression, leaving the term "high full employment" for real full employment. The Phillips curve was a refinement of this idea —suggesting that one can do business with the dragon—buying some reduction in the degree of inflation by feeding him a certain number of jobs (I almost said a certain number of unemployed, but that would make the metaphor too vivid), seeking the optimum position on the curve where the gain from this dragon trade is maximized.

There is a tendency to suppose that the optimum position is where the number of unfilled jobs is equal to the number of unemployed workers. This assumes that an unmanned job is somehow the equivalent counterpart of a jobless man, but a jobless man is a human tragedy, while an unfilled job is merely an indication that output could be greater if more productive resources were available; i.e., that output is limited only by the availability of productive resources and not by an insufficiency of demand. This would call for congratulation rather than for complaint. The complaint, when contrasted with concern about jobless men, only reflects the fear that unfilled jobs lead to inflation. The real trade-off is between employment and inflation.

Since it is marginal equivalence which gives us an optimum, we should strive for the point between low and high full employment where the harm from additional inflation is just equal to the good from the accompanying increase in employment.

To recognize such an optimum we must first decide what would be

acceptable terms of trade—how much additional inflation would be worth suffering for the sake of a 1 percent increase in employment. In the range between 2 percent and 6 percent unemployment it seems that a 1 percent increase in employment carries with it about a 2 or 3 percent increase in the national product. This is a measure of the economic gain.

The distributive harm from having an inflation higher by 1 percent per annum consists of the transfer of income from those who buy to those who sell at the higher prices (including the sale of their services for income). This transfer amounts to 1 percent of the national product (during the first two years).

In a hyperinflation there is a further loss from disorganization of cost and profit calculations. But an economy that had become accustomed to a considerable rate of inflation—even as great as 25 percent per annum—would have developed automatic and regular adjustments to wages, rents, etc., and a correspondingly high money rate of interest. The disorganization would not be very serious, or in any case not very different as between a 24 percent and a 25 percent rate of inflation. These adjustments would not correct the urge to minimize cash balances, which involves a wasteful shift of time and energy from production (or from enjoying consumption) into more frequent shopping. The remedy for this—a subsidy to holders of cash to bring the private cost of liquidity closer to the zero social cost of providing it—is probably too unorthodox to be applied, but this element of damage from the rate of inflation being greater by 1 percent per annum would clearly be minor.

The first thing to consider in estimating the distributive harm—the harm from the transfer of 1 percent of the national product from buyers to sellers—is that most members of society are both buyers and sellers. In an inflation people pay higher prices but they also earn more money. In an economy which has adjusted itself to a continuing inflation, a very large part of the transfer will cancel out in this way, and in the model of ideal adjustment there is complete cancellation and no transfer.

But in an unprepared-for increase in inflation in the real world there will be considerable transfer to others from those whose selling prices have risen less than their buying prices—basically those with incomes fixed in money terms like pensions, rents, interest, and the salaries of nonmilitant employees like university professors. But these could not cover more than say a quarter of the total payments and therefore could not amount to more than one quarter of 1 percent of the national product.

While the greater part of these redistributions cannot be disentan-

gled from the large element of chance that pervades our economy in any case, some may be welcomed and others deplored. There will, however, be serious injustices to recipients of modest fixed pensions, alimony, relief, or health insurance payments, or interest on modest savings. Where these cannot be corrected directly by collection from the seller-beneficiary, they can (and should) be compensated out of the general funds of society. Such compensation could not amount to more than a quarter of the quarter of 1 percent of the national product which covered all the uncanceled payments, or one-sixteenth of 1 percent of the national product.

The second point to remember is that compensation is not a social loss, but only a transfer to the victims from those who have to pay additional taxes raised, or suffer the reduction in benefits from government expenditures curtailed, to provide the compensation money. Much of this would even be a transfer from those who have gained in the inflation what the victims have lost. In any case, the only real loss would be from the increase in administration in collecting taxes and distributing the compensations and from the disturbance to the price mechanism by the additional taxes. We may estimate this loss at not more than one-quarter of the compensation or one-sixty-fourth of 1 percent of the national product. This is the figure that should be compared with the 2 or 3 percent of the national product gained from a 1 percent increase in employment. The acceptable trade-off ratio is therefore more than 100:1. One-tenth of 1 percent increase in employment more than cancels the harm from 10 percent inflation! The moral is that the optimum point must be very close to high full employment.

I would apologize for so much elementary arithmetic were it not that many policy suggestions seem to be based on the assumption that 1 percent of additional inflation is worse than 1 percent of additional unemployment. Proposals to restrict effective demand so as to guarantee price stability imply that inflation is an absolute evil, so that 1 percent of additional inflation is worse than any degree of unemployment; while those who would restrict effective demand to the point where wages would be stabilized and increased productivity showed itself in falling prices, must subscribe to an arithmetic that shows zero inflation as infinitely worse than any degree of unemployment.

Such weird arithmetics are usually obscured by colorful pronouncements such as that even incipient price increases show that the economy is "overheated," or that a high level of unemployment constitutes a boom that is bound to be punished by a "bust"—metaphors that suggest either the perils of primitive steam engines or the angry fates of Greek tragedy. Fortunately practice has not followed these preachings since the Great Depression. This is because unemployment is very un-

popular and political pressures and the desire for reelection cause the authorities almost always to fail in their declared intention to prevent inflation at all costs. But compromise policies have adopted points close to low full employment where an additional 1 percent of unemployment is accepted to avoid less than 1 percent of additional inflation.

There is, of course, also concern about the balance of payments, and for this a decrease in employment is just as good as or even better than a checking of the inflation. Whereas price disinflation can only gradually build up more export trade, an increase in domestic unemployment brings about an immediate and reliable improvement by a reduction in the quantity of imports we can afford to buy.

This is not the place to discuss sacrifices that many countries are making for the sake of an imaginary "stability of the exchange rates" demanded by the international gold standard and its successors. But I cannot refrain from pointing out that they have developed a more friendly face for economic warfare. Instead of "exporting unemployment" by each restricting others' exports, they peacefully increase their own unemployment. That this results in a decrease of imports, which happen to be the other countries' exports is just tough luck. "Beggar my neighbor" has been replaced by "beggar myself."

A parallel proposal in the political field was made a few years ago to save the United States and the Soviet Union the great expense and the dangerous uncertainties of developing and maintaining effective Intercontinental Ballistic Missiles. It was proposed that H-bombs be buried in the principal cities of both countries, and that each government be given by the other a reliable electronic capability of detonating the bombs in the cities of the other country. Perhaps it is an indication of the lagging of the other social sciences behind economics that the proposal was dismissed as "too logical."

There are yet other arguments brought against moving to the optimum trade-off point near high full employment. The structural unemployment argument claimed that even when we had 5 to 6 percent unemployed, we were very close to full employment, the unemployed having been made unemployable by automation. The evidence for this —a negative correlation between years of education and percentage unemployed—may, however, only show that counting the years of schooling is an easy way of deciding how to allocate scarce jobs.

Then there is the fear of "accumulated" inflation. A 1 percent increase in the rate of inflation will make prices higher (than they otherwise would have been) by an average of one-half of 1 percent during the first year, and by 1½ percent during the second year (hence the

1 percent transfer from buyer to seller in the first two years). In the following years it will be 2½ percent, 3½ percent, etc., above the previous trend line. Soon the compound interest factor becomes significant and it might be thought that with such a growing shift of income from buyers to sellers the burden of compensation would become intolerable. But if the inflation continues at the higher rate the whole necessity of compensation fades away. As the economy gets used to the inflation, it develops the appropriate automatic adjustments of wages, rents, pensions, etc., between the buyers and sellers so that the need for public compensation diminishes rather than increases.

The related consideration of the danger of disorganization is much more cogent. The markups may also adjust to the inflation. Larger and larger markups may be called for as the adjustments which correct the inequities to the buyers simultaneously neutralize the benefits to the sellers from the previous markups. It might then be rational to suffer the current loss from a low level of employment as the price of preventing hyperinflation in the future.

I would now give less weight to this consideration than I once did. Inflation may not accelerate. We have seen inflations continuing for a long time at about the same rate and even slowing down. But still there is the danger.

The economic cost of holding the economy down to low full employment so as to avoid this danger—some $50 billions per annum—seems far too large an insurance premium against the risk of sliding into hyperinflation in the future, especially since the payment, instead of guaranteeing our future wisdom, would even increase the danger if popular resentment at the unemployment should bring support to irresponsible inflationists.

But to plead for a more rational choice of the optimum point on a given unemployment-inflation spectrum is only a secondary and interim issue. My primary purpose is to argue that we do not have to accept the spectrum. We can arrange to enjoy the benefits of high full employment and price stability.

I am referring of course to incomes policy and the wage and price guideposts. These have not been too successful up till now, but my thesis is that greater success will follow a better understanding of the guideposts, not as a substitute for monetary and fiscal regulation of effective demand, or even as a parallel support to these, but as a replacement for a part of the competitive price mechanism that has broken down (or perhaps never worked properly).

In a strictly market governed economy, i.e., where prices always rise or fall so as to make supply equal to demand, full employment does

not require any monetary or fiscal policy. We are in a neoclassical world in which price levels adjust the real value of any nominal money supply to fit the demand for real cash balances.

Monopolies and monopsonies do not interfere with this. They only raise the prices obtained by the monopolists, or lower those paid by the monopsonists, relatively to other prices, thereby raising their real incomes relatively to the incomes of others.

But if we have a "double monopoly" when monopolists sell to each other, no others are there to be exploited, so that both together cannot gain anything. As they raise their selling prices they see their buying prices rise, and while their margins rise in absolute terms these cannot rise relatively to each other and a price race ensues. If the authorities do not increase the total rate of spending, there is a deficiency of demand. But the prices, instead of being pushed down by the market, continue to be raised by the monopolists. Sales and output fall as much as may be required to induce the monopolists to give up the race. This is the point of low full employment.

With a double monopsony instead of a double monopoly, all this is in reverse. Each reduces his buying price to increase his share and is frustrated by the fall in his selling price which is the other monopsonist's buying price. The margins fall absolutely but not relatively, and unless the authorities reduce the level of spending there is excess demand. Each wants to buy more than is available. But the prices, instead of being bid up by the market, continue to be pushed down by the monopsonists until the pressure of excess demand is great enough to induce them to give up the downward race and they merely prevent prices from rising.

This condition is familiar to us as "repressed inflation" (when it is usually the government who keeps prices down). The opposite case, at low full employment, when sellers would like to sell more but the prices do not come down, is, however, not called repressed deflation but recession or depression. Perhaps depression is a compression of the two words repressed deflation. Our problem is one of "overrepressed" deflation. The monopolists do not merely stop the deflation from reducing prices; they continue to raise them. Or perhaps it is better to put this the other way round. There is not enough deflation (and unemployment) to stop the double monopoly completely from raising prices. We are above "low full employment."

Our monopolists are the suppliers of labor power, whom I shall call labor, and the suppliers of the goods produced with the labor power, whom I shall call industry. Labor sells labor power to industry, and industry sells goods to labor. Labor and industry are of course not monopolists in the usual sense of the word, in the sense related to the

elasticity of demand for the output of the firm, but their behavior is what counts to us behavioral scientists.

Labor maintains at least a constant ratio of the wage to the cost of living (the price of goods) unless unemployment is severe. This is equivalent to a degree of monopoly or a rate of markup. (Markup is the excess of price over marginal cost, here the excess of the wage over some basic cost of living. Degree of monopoly is markup over price, rate of markup is markup over marginal cost, and the average variable cost is a proxy for marginal cost.) Industry competes vigorously, at least in its own language; but insofar as it operates with rates of markup we economists have a measure of its degree of monopoly.

The task of the guideposts is to enable the appropriate level of effective demand to yield us full employment with price stability, as in the "innocent" Keynesian theory. To achieve this the guideposts must get wages and prices to behave as if they were market determined. Fortunately we do not need to carry this to the extreme "neoclassical" degree that would make both monetary and fiscal policy unnecessary (by giving us automatic full employment via wage and price deflation). Since we can maintain adequate demand by monetary and fiscal policy it will suffice to make wages and prices behave as if market determined only within the narrow range that is compatible with our other objective—price stability. But even for this we must overcome the double monopoly.

This has of course been recognized by the administrators of the guideposts, but an overemphasis on average wage and price levels has led to the neglect of the relations among the different wage rates and among the different prices.

The emphasis was on stabilizing the price level by getting wages to rise in proportion to the increase in productivity. This is a loose expression generally taken to mean the average product of labor. But what is necessary for price level stability for a given rate of industry markup is that marginal costs should on the average remain stable, and that means that wages should rise with the average of the marginal productivities of labor.

The formula that wages must rise in proportion to productivity was naturally seized upon by workers in industries (or even firms) with a large increase in average product to claim a correspondingly large wage increase. The misinterpretation is reminiscent of the equally disingenuous way Colin Clark's taxable capacity guidepost—that raising in taxes more than 25 percent of the national income leads to inflation —was perverted by businessmen into a warning against taxing any individual more than 25 percent. Like the businessmen's perversion, the perversion here was also properly denounced. But the correction usual-

ly took the form of insisting that the proper criterion for the wage increase is the increase in productivity not of the industry but of the economy as a whole.

This is the wrong correction. It does not indicate what the wage in the industry should be because it is the guidepost only for the average wage in the whole economy, and as we all know the average does not really exist. It is only a figment of the arithmetic. The wrong correction was taken to mean that no wage increase that exceeds the national average increase in productivity is in conformity with the guideposts and that any wage increase that does not exceed this average is in conformity. But this in strict logic implies that price level stability requires an absolute freeze in relative wages. The case is the exact opposite of the famous hotel porter, who when given what he had said was his average tip, gratefully told the guest that it was the first time he got as much as the average.

Guideposts for particular wage rates cannot be developed without considering the functions of differences in relative wages. One of these is to induce workers to go where they are needed more. Bringing this in makes it possible to give a reasoned answer to the workers in the high productivity gain industries. They can then be shown that if the reduction in the price of the industry's product made possible by the high productivity gain leads to an increase in demand greater than the increase in output per man, then more workers should be drawn into that industry from other industries and an above-average wage increase is appropriate. But if the increase in demand is less than the increase in output per man, then workers have to be drawn away from this industry into others, and the wage increase, if any, must be below average.

All this comes out even more clearly if we look at industries with below-average increases in productivity. If the wage rises at the guidepost rate for the average wage, the price of the product will have to be raised by the increase in the marginal cost, which is equal to the excess of the wage increase over the low increase in the marginal product. If with this price increase the demand falls or increases less than output per man, labor must be shifted from this industry to others, and a below-average particular wage increase is indicated. But if the product is a noninferior good on which a larger part of income tends to be spent as total income increases, and this tendency more than overcomes the rise in price so that there is an increase in demand, and the increase in demand is greater than the low increase in output per man (a combination that is quite normal, especially in the growing service industries), then more workers should be drawn into this industry, and

for this purpose the wage must rise more than the average although the increase in productivity is less than the average.

Clearly what is creeping in is supply and demand, or rather the need for an artificial replacement for the malfunctioning market organ disorganized by the double monopoly.

The guideposts for particular wage rates must be expressed as rules governing departures from the central guidepost for the average wage. These departures can be based only on indicators of the conditions of supply and demand for labor in each labor market. Productivity gains can influence the guideposts only via their effect on supply and demand in the labor market. The wage increase should be greater than the average where the demand is greater relative to the supply than on the average, and it should be less where the reverse is the case. The departures must furthermore be such as will keep the average of all the wage increases in conformity with the central guidepost. Insofar as the upward pressures of the double monopoly can be relied on, the guideposts may be expressed as maximum increases. If we had double monopsony, they would be better expressed as minima. But they are really the desirable changes, and departures from them are no better in one direction than the other.

Until such rules are worked out and applied at least to the important labor markets, the guideposts can have only limited and temporary influence and are liable to be disobeyed, discredited, and destroyed as the pressures and temperatures of supply and demand inequalities accumulate beneath the procrustean freeze of relative wages.

This metaphor is not as mixed as it sounds. The guidepost practitioners, not unaware of the issues I have raised, have in fact been depending on unofficial departures from the guideposts, in both directions, in response to supply and demand pressures, to let some of the accumulated steam escape—thus thawing some of the freeze. But such departures are more likely to reflect organized power of pressure groups and relative lack of social responsibility rather than the social need for moving people between the different labor markets.

They have also depended on the mobility of workers in response to the relative ease of finding employment in different labor markets. But this is an effective adjustment mechanism only when there is severe unemployment and men are trying hard to find any job whatever. And even if there were such high mobility in response to the relative ease of finding jobs, a frozen set of relative wage rates would result in relatively large numbers of unemployed workers in the markets where the wages and the job conditions are relatively attractive, waiting and hoping to be hired, while the markets where conditions are relatively unat-

tractive would be frozen with a scarcity of workers, a small output, high prices, and enormous profits protected from competition. In short, the prices and wages would fail to perform their social functions.

The reliance on such very poor substitutes for a real set of guide-posts (which itself would only be a poor substitute for the missing piece of market mechanism) seems to me to be due less to the difficulties of developing the rules than to a powerful series of objections that would certainly be raised and the political difficulties that would emerge.

Perhaps the most powerful objection is the cry of "price control." This is not a logical objection, because price control prevents, or tries to prevent, a price from increasing when demand exceeds supply. It compels the price to disobey the market which is telling it to rise. The particular guideposts would do just the opposite. They would be reflecting the market forces of supply and demand and telling the wage to obey these—to rise more where there is a relatively greater excess of demand and to rise less where there is a relatively greater excess of supply—instead of serving the double monopoly. The guideposts would be interfering not with the market mechanism but only with the interferers with the market mechanism.

Another powerful objection is the cry that collective bargaining would be weakened by losing its primary task: the setting of wage rates. This is a more logical objection and no argument can shake those for whom collective bargaining has become an absolute good—an end in itself. But those who view collective bargaining only as an instrument for improving the condition of the people could perhaps be made to see that the guideposts would serve this purpose much better where the collective bargaining of a trade union has become one side of the double monopoly.

Closely related to this is the cry that government regulation of wages and prices constitutes a totalitarian denial of individual liberty. It is not easy to get the singers of such slogans to look and see if anybody's liberty to work or buy or sell is really being interfered with, or indeed any liberty other than the liberty of the double monopoly to hamstring the price mechanism.

Subject to severe criticism would be the apparent concentration on the regulation of wages, with the regulation of prices all but forgotten. This apparent one-sidedness is inevitable because average wages have to rise with average marginal productivity if average prices are to be stable. This means that there must be enough price reduction to balance the increases; otherwise the employer monopolists would be taking advantage of the wage regulation to increase their relative markup —an injustice which if not promptly removed would certainly destroy

the whole guidepost attempt. The price guideposts cannot, however, be based on the relatively simple indicators of supply-demand relationships possible for wages—such as the percentage of unemployed or of unfilled vacancies. In general there will have to be reliance on competition to prevent any general increase in the rate of markup with investigation of complaints of prices rising more than costs or failing to fall as much as costs. This is the general problem of antitrust or antimonopoly with its various measures for increasing competition by outlawing restraints of trade, reducing or removing protective tariffs, unloading government stockpiles, applying more direct government influence as a large purchaser, and many other such devices, including as a last resort a greater activation of competition or threats of competition by government "yardsticks" or by government "counterspeculation" wherever competitors are being discouraged in spite of apparently high profit opportunities.

In this I would give more credit than I once did to the effectiveness of sermons by the administration, perhaps accompanied by hints of other measures up its sleeve to deal with sinning monopolists. The denunciation of a proposed steel price increase by President Kennedy seems to have been a prime factor in raising the level of low full employment.

In the regulation of prices and wages there has to be a symbiosis of legal with moral pressures. Laws which are not accepted as morally valid are not successful—as we have seen with prohibition—but the enforcement of appropriate laws can help to establish the customs from which the moral support can develop—as we are seeing in some fields of racial integration. In the matter of guideposts, the development of public opinion pressure can strengthen and in turn be strengthened by the use of sanctions to develop habits of compliance; and the crude wage regulations and competitive stimulants can grow into a more and more efficient substitute for the missing piece of the price mechanism.

We must remember that this is the way in which the familiar basic institutions of the free market were evolved. Richard T. Ely, for whom this lecture is named, first taught me that markets and money, prices and contracts did not drop from heaven but grew out of painstaking and daring innovation and experiment.

A sanction frequently proposed, both as a means of strengthening the guideposts and as an alternative to them, is for the authorities to declare that though they will provide all the effective demand needed for full employment, they will sternly refuse to provide any more under any circumstances, so that if prices rise there will be insufficient demand for full employment.

The argument is that if this threat is believed, both labor and indus-

try will obey the guideposts, or refrain even in the absence of guide-
posts from raising wages and prices to levels that would result in
inflation and unemployment.

This is not good advice because even if the unemployment succeeded
in containing the inflationary pressures, these would only emerge with
greater force whenever an attempt was made to permit higher employ-
ment again in the future. Second, the authorities would never be able
to carry out the promised cruel and unusual punishment, and this
being foreseen, the threat would not be believed. Third, even if it were
believed, each particular trade union or industrial monopoly would feel
that the future depression was coming anyway because of the wicked-
ness of everybody else, so that its own restraint would only increase
the ill-gotten gains of the others. And finally the use of unemployment
or the threat of unemployment to fight inflation may work the wrong
way. Just as in foreign policy, it is unsafe to assume that while we will
agree to negotiate only from strength, the enemy, unlike ourselves, can
be pressured into reasonableness. We may find him as determined as
we are.

However, if the authorities made good their threats and remained in
power and gained credence for their threat for the next time the low
employment threshold was crossed, this would strengthen support for
the development of sanctions for more complete guideposts to force·
everybody to behave in ways that would be better for all.

A particularly interesting variant of the "stern refusal" policy was
recently provided by Israel. A sharp increase in concern for the bal-
ance-of-payments position led the Minister of Trade and Industry to
propose a uniform subsidy on all exports, to be financed by a uniform
tax on all imports. This was opposed by the Minister of Finance, who
proposed instead to give subsidies only where really needed. A closer
examination showed that the counterproposal was based not on eco-
nomic illiteracy but on a recognition that a partial devaluation would
quickly become a complete one, which would soon thereafter be
nullified by a wage increase induced by "unneeded" profits. Behind
this was the further calculation that it would take a long time before
the bureaucracy would negotiate the discriminatory subsidies. Mean-
while economic recession, created by restrictive monetary and fiscal
measures and unalleviated by any significant growth of exports,·could
greatly increase the efficiency of the economy by inducing labor to give
up many featherbedding, job-protecting practices like "last in, first
out" in jobs (which prevented useless workers from being fired). The
leaders of the Association of Manufacturers were persuaded to support
the counterproposal by astonishment at the labor organizations' agree-

ing for the first time to waive an increase in cost-of-living allowances and hope of being able to negotiate export subsidies for their own businesses.

The benefit from depreciation probably would be shortlived, there being no effective guidepost system, but it seems to me that continued depressions would strengthen rather than weaken the cost-increasing (and thereby employment-reducing) job-protection practices and do serious longer-term harm to the economy by reducing investment and the technical progress that goes with it. Nevertheless I sadly expect to see this experiment repeated in other countries. The British price-cool-off and wage-freeze seems to me to be one such case.

On the other hand persistent and successful maintenance of high full employment would yield a long-term benefit in reducing the average rate of markup. There would in the first place emerge high profit rates. The markups, adjusted by competition to yield normal rates on the average over good times and bad times, would yield superior profits on the larger outputs of the continuing good times. We can see this result in the sharp rise in profits during the recent increase in the level of output; and you must all have been getting literature from the labor unions which plays up this result of higher output as the root cause of inflation. But whatever we may think of its economic analysis, this literature reflects a pressure for higher wages in response to the higher profit rates.

Any increase in money wage rates induced by such profits by themselves only result in further increases in prices and a sharpening of the double monopoly inflationary pressure. But the high profits are visible also to entrepreneurs and would-be entrepreneurs. As they acquire confidence in continuing good times, some seek to exploit the prospects of adequate profit rates from lower markups. Such optimistic new competitors now escape the normal decimation that was their fate in the past during the recurring bad times, and their continuing competition brings about a reduction in the average rate of markup, lowering profits to normal. This gives labor a larger share of the total product (accompanied by false claims to credit for this by the organizers of pressure for higher money wages).

In the longer period, a more complete adjustment to continuing prosperity permits both real wages and profits to be raised by a reduction of expenditure on relatively inefficient but cheaper standby equipment for use only at peak output periods and by easier maintenance of continuous technological research. A greater supply of the less demanding kind of entrepreneurship that does not have to worry about depressions or recessions leads to a further, and this time parallel, reduction

in both markups and profit rates. This increases the relative as well as the absolute share of labor and permits the average wage to rise more than the average marginal product.

We have a different natural hisory if administered inflations are combated by administered recessions. The short period effect of a disinflationary cut in demand is a cut in profits, from a fall in sales followed by distress price cuts. When adjustment had been made to the lower level of demand, the bankruptcy of weaker competitors permits somewhat higher markups to restore almost normal rates of profit with lower sales. And in the long run both markup and profit rate would return to the old norm.

We may speculate briefly on the general shape of our future in the absence of a more complete guidepost system. As the economy becomes more affluent, it is able to maintain an increasing proportion of unemployed at rising standards of normal decency. This increase in the "demand for unemployed" is accompanied by a parallel increase in the supply. Labor, scared by the growing unemployment, concentrates more and more on protecting jobs against technological progress, while the affluence makes them increasingly able to sabotage "automation" while demanding higher pay for fewer hours. This lowers the level of low full employment. It takes a larger percentage of unemployment to keep inflation down to any specific rate.

But how important is it that we succeed, by a completed guidepost system, to reach and maintain high full employment without inflation?

There is first the direct increase in the national product from having some 3 percent more of the labor force at work. This would mean a bonus of 6 to 9 percent of national product or some $50 billion worth of additional output. The bonus is the gap between low full employment national product and high full employment national product—an increasing fraction of the growing national income if, as I shall argue, the low full employment level (percent employed) is likely to fall over time. These figures are debatable, but even if they came to $100 billion, the mere increase in output would not add much to the happiness or well-being of the people of the United States. A Parkinsonian law of Veblenian conspicuous consumption makes men's urgent needs grow to absorb the available means to satisfy them. Perhaps the most significant direct result of greater growth and greater consumption at current American levels (of both affluence and poverty) is only the more rapid exhaustion of the mineral resources of the planet.

The second and much more important result is the indirect effect on the image of the West in the eyes of the challenging East and the uncommitted South. It would remove the remnants of belief in the Marx-

ian theory that capitalism is inferior to communism because it must maintain a "reserve army of unemployed" to keep the exploited workers in increasing misery. It does not matter that the argument has been transmogrified to say that a totalitarian society is superior to a democratic society because it alone is able to regulate wages and prices for full employment without inflation. A successful guidepost system would vaporize the grain of truth in this, demonstrate the combination of these two objectives with two other great objectives, greater efficiency and freedom, strengthen (or perhaps one should say rescue) the esteem of the South for the West, and weaken the belief of the East, and even of the Further East, that the West is hypnotically moving toward its Marx-ordained doom. It would be an important step in the integration of order with freedom, which with the slow and intermittent growth of freedom within the order of the East constitutes the convergence on which the future of civilization depends.

The third result, possibly even more important than the second, is the social hygiene of full employment. The list of important social benefits is so long that it strains the credulity—but only until one sees them as the fruits of the removal of a miasma in which a multitude of diseases flourish. Among these diseases are the feelings of uselessness and failure, the unsatisfied need of being needed, with all the resentments and the further injustices that are set in motion by these frustrations, the fear of and resistance to innovation as a destroyer, the concentration on monopolistic, legislative, and other protections against competition, the unjust condemnation of the unemployed as unemployables, and worse still the conversion of some of them into genuine unemployables. Most important of all is the vicious and apparently inevitable concentration of the unemployment on underprivileged and alienated minorities, fostering undiscriminating hatreds as these develop their own subcultures of poverty and race in which they see themselves as victims of the "system" which owes them a living without work until they can get employment at fair rates of pay, which comes to mean the rates of pay of workers with skills and experience they do not share.

The fourth and crowning result is the way the first three cooperate with and magnify each other. The alleviation of the unemployment-nurtured diseases increases potential output, improves our image in the South, and fosters the acceptability in the East of coexistence over collision. The greater output makes possible an easier carrying out of the tasks, imposed on us by both morality and prudence, of helping the South break out of its poverty, and rehabilitating the domestic minorities so badly damaged by past injustice. Meanwhile the improvement

MILITARY MANPOWER PROCUREMENT

THE SUPPLY OF MILITARY PERSONNEL IN THE ABSENCE OF A DRAFT*

By STUART H. ALTMAN, *Brown University*
and ALAN E. FECHTER, *Institute for Defense Analyses*

In April, 1964, President Johnson directed the Department of Defense to undertake a comprehensive review of our military manpower procurement system and to determine whether it would be feasible to shift from the current partial reliance on conscription to a system in which manpower needs would be met entirely by volunteers. Part of this review involved estimating the budgetary costs (as opposed to social costs) of shifting to an all-volunteer procurement system. This paper summarizes the method used in estimating these costs for active duty enlisted and officer personnel, exclusive of doctors and dentists.

The Supply of New Recruits with a Draft and No Change in Relative Military Pay

To understand the way in which the market for military manpower operates, we examined actual recruiting experience for enlistees and officers since 1954. We discovered that there were many legal and administrative constraints on recruiting practices. These made it difficult to separate variations in the number of new enlistees or officers recruited caused by supply factors from those caused by demand factors. In short, there were serious identification problems involved in a supply analysis based on total recruitment data [1] [2]. Because of these problems, we had to limit our supply analysis to groups who were procured through programs that were relatively free from these constraints: Army enlistees in Mental Groups I-III[1] and officers commissioned through ROTC programs in "voluntary" schools.[2]

[1] Various tests have been designed to measure the ability of new recruits to absorb military training. The Armed Forces Qualification Test (AFQT) is the test used by all services to classify enlisted applicants. On the basis of AFQT scores, applicants are grouped into one of five mental categories, representing a regressive range in mental ability, from very rapid learners (Mental Group I) to very slow learners (Mental Group V). All Mental Group V applicants are by law excluded from military service. In addition, the services have, at times when there were excess supplies of enlistment applicants, severely restricted enlistments from lower Mental Group IV.

[2] Some colleges and universities require all physically qualified nonveteran male freshmen and sophomores to participate in ROTC. Other institutions offer the ROTC program, but

It was found that the following variables were important determinants of the level of enlistments, given the existence of military conscription: (1) the size of the 18–19 year old male population; (2) the male 18–19 year old unemployment rate; (3) major changes in draft pressure; and (4) seasonal factors. These variables were incorporated into the following regression equation, estimated from quarterly data for the period July, 1956, to June, 1965:[3]

$$\frac{E}{P} = .03018 + .35670\ U^* + .02007\ D_H\dagger - .01341\ D_L - .01024\ S_p \quad (1)$$
$$\phantom{\frac{E}{P} = }(.07696) \qquad (.00865) \qquad\quad (.00703) \qquad (.00522)$$
$$+ .02107\ S_u^* - .00517\ F$$
$$(.00620) \quad\ (.00588)$$
$$R^2 = .67$$

NOTE: Figures in parenthesis denote standard errors.
  * Significant at the .01 level.
  † Significant at the .05 level.

where,

E/P = quarterly Army Mental Group I–III enlistments divided by the 18–19 year old male population.

U = the 18–19 year old male unemployment rate.

D_H = a dummy variable for periods of high draft pressure (the four quarters during the Berlin buildup crisis).

D_L = a dummy variable for periods of low draft pressure (the four quarters of FY 1965 during which time draft calls were low and there was extensive publicity about the long-run possibility of eliminating the draft).

S_p = a spring seasonal dummy variable.

S_u = a summer seasonal dummy variable.

F = a fall seasonal dummy variable.

The coefficients of equation (1) support the following conclusions: (1) periods of relatively high civilian unemployment tended to produce greater than average numbers of volunteers; (2) periods of relatively high "draft pressure" (increased world tensions) also produced higher than average levels of volunteers, while periods of low "draft pressure" produced less than average enlistments; (3) enlistments were above average in the quarter just following high school graduation and below the yearly average in the fall and spring.

---

leave the decision about enrollment in the hands of the student. We have designated the former schools "compulsory" and the latter "voluntary."

[3] Although military earnings was found to be an important factor in enlistment behavior, its effects could not be estimated directly from time series data because of insufficient statistical variability over the period studied. Therefore, no explicit variable was included in equation (1) to account for it. However, estimates of the effects of military earnings on enlisted and officer accessions based on cross-section data are discussed later in this paper.

Enlistments with a draft were projected for the years 1970 through 1975 assuming that lower draft calls in the future would produce a level of draft pressure similar to that existing in FY 1965. Estimated future enlistments were based on projected 18–19 year old male population[4] and were computed from equation (1) for two unemployment levels. The first corresponded to the male 18–19 year old unemployment rate in FY 1957 (11.5 percent) when overall unemployment was at the relatively full employment rate of 4.0 percent; the second, to the average 18–19 year old rate between FY 1957 and FY 1965 (14.9 percent), a period when the overall rate averaged 5.5 percent (Table 1).

TABLE 1
PROJECTED SUPPLY OF ENLISTED VOLUNTEERS WITH A DRAFT,* FY 1970–75
(In Thousands)

Years	18-19 Year Old Male Population	5.5% Unemployment Level			4.0% Unemployment Level		
		Total	Army	Other Services†	Total	Army	Other Services†
1970–71	3,875	414	150	264	342	124	218
1972–73	4,090	444	161	283	375	136	239
1974–75	4,255	458	166	292	387	140	247

* Includes nonprior and prior service enlistments plus Navy two-year active duty Reserve enlistments.
† Represents enlistment capabilities including 10 percent Mental Group IV personnel. In some years the estimated potential supply was somewhat below requirements. We assumed that, with a continued draft, these services would make some changes in recruitment practices so that they would not have to resort to the draft to make up their deficits.

Since equation (1) was estimated for Army Mental Group I-III enlistments only, several additional assumptions were made:

1. The Mental Group I-III enlistments of the other services would grow at the same rate as Army MG I-III enlistments. Fiscal year 1965 was used as the base.[5]

2. Mental Group IV accessions would be limited to 20 percent of total new Army enlistments and, if necessary, 10 percent of new enlistments of the other services.

3. Prior service personnel would continue to return to active duty in about the same numbers as in the recent past.

In recent years almost all new officer accessions have been college graduates. With the draft acting as a stimulus, more potential officers have been seeking commissions than otherwise would, and a buyers

[4] Projections of population were obtained from the U.S. Bureau of the Census.
[5] FY 1965 was a year in which each service had difficulty meeting strength objectives. Therefore, it was unlikely that quota controls restrained Mental Group I-III enlistments for any service during that period.

market has existed. The limited number of available positions has been allocated by raising educational standards.

An excess supply in many of the officer-procurement programs has tended to hide the effects of population on officer supply.[6] However, there did not appear to be any significant excess supply of participants in first year ROTC. The "enrollment rate" in ROTC-I in "voluntary" schools proved to be extremely stable except for years in which major changes occurred in either draft policy or policy regarding length of initial tour of active duty. In general, whenever a policy was implemented that tended to reduce the likelihood of a potential ROTC student being drafted or resulted in his having to serve a longer tour of obligated duty, the enrollment rate fell [2]. We concluded that, given draft policies and

TABLE 2

. PROJECTED POTENTIAL SUPPLY OF NEW OFFICERS WITH A DRAFT* BY SERVICE,
FISCAL YEARS, 1970–75
(In Thousands)

Fiscal Year	Army	Navy	Marine Corps	Air Force	Total
1970–71	19,.0	11.4	2.7	13.3	46.4
1972–73	19.2	11.8	2.8	14.0	47.8
1974–75	19.5	12.2	2.9	14.3	48.9

* Projected accessions were estimated on the basis of the relationship between officer production for 1964 and initial inputs into training for the ROTC and officer candidate programs and on the basis of service projections of planned output for the other commissioning programs. Attrition rates in the ROTC and officer candidate programs were assumed stable and inputs were expanded equiproportionately in relation to the growth of the population from which the programs draw their applications.

a stable length of initial tour of duty, voluntary enrollments in ROTC would vary proportionately with variations in enrollments of under-graduate male freshmen.

In projecting the supply of new officers with a continued draft, we assumed that commissioning programs which have had an excess supply of applicants would not be affected by population changes but would change in accordance with service procurement aims. This limited programs affected by population to ROTC and officer candidate.[7] Population growth within these programs plus planned changes in other procurement programs were estimated to provide the services with a potential supply which will more than meet new officer requirements at pre-Vietnam strength levels (Table 2).

[6] In these programs, population would not affect officer accessions; it might affect the supply of applicants to these programs if recruiting effort were stable. It appears, however, that, given program size, recruiting effort varied inversely with population, causing the number of applicants to remain fairly stable.

[7] Officer accessions from ROTC and related programs were projected from Office of Education projections of male freshmen enrollment statistics lagged four years. Projected accessions from officer-candidate programs were based on Office of Education projections of the number of males graduating from college. Details are contained in [2].

The Effect of Conscription on the Supply
of Military Personnel

If the draft were eliminated and other factors affecting supply did not change, the services would lose not only all their inductees, but the number of new recruits who would voluntarily supply their services to the military would also fall substantially. The existence of the draft has

TABLE 3

RESPONSES OF REGULAR ENLISTED PERSONNEL ON THEIR FIRST TOUR OF
ACTIVE DUTY TO "WOULD YOU HAVE ENLISTED IF THERE HAD BEEN
NO DRAFT?"*—BY SELECTED CHARACTERISTICS
(Percent Distribution)

Characteristics	Total	Yes, Definitely	Yes, Probably	No, Probably	No, Definitely	No Idea	Draft Motivated†
Total.........	100	30.1	27.6	20.1	15.2	7.0	38.0
Service							
Army.........	100	27.1	25.6	20.7	19.4	7.2	43.2
Navy.........	100	32.3	30.2	17.9	12.3	7.3	32.6
Air Force......	100	26.6	26.4	23.8	16.0	7.2	42.9
Marine Corps..	100	37.0	28.7	17.4	11.3	5.6	30.4
Age at enlistment							
17 to 19 years..	100	34.2	29.5	17.2	12.0	7.1	31.4
20 to 25 years..	100	17.6	21.6	29.1	24.9	6.8	57.9
Education‡							
Less than high school graduate	100	42.9	28.8	11.7	9.7	6.9	23.0
High school graduate.......	100	27.2	28.2	22.4	14.9	7.3	40.2
Some college or more.......	100	16.3	23.1	28.3	26.0	6.3	58.0
Mental Group							
Groups I and II	100	25.1	27.4	22.8	18.5	6.1	44.0
Group III.....	100	33.9	28.0	18.1	12.6	7.5	33.2
Group IV.....	100	36.6	27.9	14.8	11.8	8.9	29.2

* The question asked, "If there had been no draft and you had not had any military obligation at the time you first entered Active Military Service, do you think you would have entered the service?"
† Number answering "No probably" or "No definitely" as a percent of the total, excluding "No idea."
‡ At time of entry on active military service.
SOURCE: Department of Defense Survey of Active Duty Military Personnel, as of Oct., 1964.

prompted many to enlist or seek commissions rather than to serve as draftees. These "draft-motivated" volunteers could not be expected to enter service if the draft were eliminated. To determine the extent to which the existence of the draft has influenced the level of voluntary enlistments and officer accessions, a sample of active duty personnel were asked specifically whether they would have entered the service if there had been no draft and they had no military obligation. The re-

sponses to this question for enlisted and officer personnel serving their first obligated tour of duty are summarized in Tables 3 and 4.[8]

Of those responding, 38 percent of the enlisted men and 41 percent of the officers indicated that they "definitely" or "probably" would not

TABLE 4

RESPONSES OF OFFICERS ON THEIR FIRST TOUR OF ACTIVE DUTY:
"WOULD YOU HAVE ENTERED SERVICE IF THERE HAD BEEN NO DRAFT?"—
BY SERVICE AND SOURCE OF COMMISSION
(Percent Distribution)

Selected Characteristics	Number in Sample*	Total	Yes, Definitely	Yes, Probably	No, Probably	No, Definitely	No Idea	Draft Motivated†
Total.....	7,662	100	36.0	20.0	24.1	15.3	4.6	41.3
Service								
Army.......	2,328	100	30.6	18.6	27.4	19.0	4.4	48.4
Navy.......	1,845	100	37.9	19.5	23.4	15.3	3.9	40.3
Air Force....	3,017	100	36.8	21.2	23.2	13.6	5.3	38.9
Marine Corps	472	100	49.8	20.3	17.4	8.5	4.0	27.0
Source of Commission								
Academy....	583	100	71.4	14.2	7.2	3.3	3.9	10.9
OCS‡.......	2,133	100	25.9	20.4	30.8	18.2	4.6	51.4
ROTC......	2,764	100	28.3	23.7	27.6	15.6	4.8	45.4
Direct appointment§	862	100	23.7	16.6	24.2	30.9	4.6	57.8
Other.......	1,320	100	60.9	16.0	13.6	5.2	4.4	19.7

* Number in sample may differ from comparable numbers in other tables because respondents who did not answer are excluded.

† Number answering "No probably" or "No definitely" as a percent of the total, excluding "No idea."

‡ This category consists of personnel who receive their commission upon completion of officers candidate school or officers training school.

§ This category consists of physicians, dentists, lawyers, and other professionals. Such personnel receive their commission through direct appointment.

SOURCE: Department of Defense Survey of Active Duty Military Personnel, as of Oct. 31, 1964.

have entered service in the absence of a draft. In both cases the highest proportion of draft-motivated personnel was recorded for the Army and the lowest for the Marine Corps. More detailed tabulations indicated that the draft was a more important factor the higher the level of educational attainment and older the recruit [3, p. 10039]. For officers, draft motivation varied significantly by source of commission. As anticipated,

[8] Because of the inherent limitations of attempts to probe human motivation by use of survey questionnaires, we attempted to check the reliability of the responses in several ways. First, the survey results were checked for internal consistency. Second, the results were compared with the responses from similar types of survey questions asked in the past. And finally, the characteristics of the draft-motivated respondents were compared with those of individuals who volunteered for active duty in greater numbers when actual draft pressure was high (as during the Berlin crisis), and who failed to volunteer when draft pressure was low (as during 1965). In general, past survey results and actual experience were consistent with the findings used in this study.

draft motivation was low for Academy graduates, and high for those commissioned from officer candidate and ROTC programs.

The supply of active duty personnel without a draft was projected for the 1970–75 period by reducing the enlisted and officer continued draft projections by the appropriate draft-motivation factors from the Department of Defense Survey. In adjusting the officer projections, we assumed that programs which offer subsidized college educations, such as service academies and regular ROTC, and prior service accession programs would not be affected by elimination of the draft.[9]

TABLE 5

RECRUITMENT DEFICITS WITHOUT A DRAFT, FY 1970–71*
(In Thousands)

	Enlisted		Officers†
	5.5% Unemployment Level	4% Unemployment Level	
Total active duty			
Annual requirements	500	512	37
Accessions—no pay increase. . .	317	272	27
Percentage increase in accessions required.	58%	88%	37%
Army			
Annual requirements.	220	232	13
Accession—no pay increase. .	106	91	9
Percentage increase in accessions required.	108%	155%	44%
Other services			
Annual requirements	280	280	24
Enlistment—no pay increase. .	211	181	18
Percentage increase in accessions required.	33%	55%	33%

* To maintain a 2.65 million active duty force.
† Excludes medical and dental officers.

A comparison of the no-draft estimates of supply with expected annual requirements to maintain active duty strength at pre-Vietnam levels indicated that it would require increases of between 58 percent and 88 percent in initial enlistments and 37 percent in officer accessions, to eliminate the recruitment deficits that would occur (Table 5). By far the largest deficit would be faced by the Army, and within each service the most severe shortages would be felt in those specialties requiring men of above average mental aptitude and education.[10]

[9] While these programs also produce some draft-motivated officers, we assumed that the excess supply of applicants to these programs was more than sufficient to offset the decline in applicants that would accompany the abolition of the draft.

[10] Lack of detailed knowledge about military production functions compelled us to make the assumption that, at least in the short run, labor would continue to be used in the same pro-

Increased Pay and Voluntary Accessions

The effect of pay on military supply was examined by analyzing the the extent to which regional variations in military relative to civilian earnings has influenced voluntary enlisted and officer accessions.

For enlisted men, the following equation was estimated from 1963 data for nine Census regions [4]:

$$\log c_i = b_1 + b_2 \log Y_i + b_3 \log U_i \tag{2}$$

$$\text{where } c_i = \left[(1) - \left(\frac{D_i \cdot E_i}{p_i} \right) \right]$$

and,

$E_i=$ 1963 enlistments in the ith region in Mental Groups I-III.

$D_i=$ the proportion of draft-motivated volunteers within Mental Groups I-III in the ith region as measured by the Department of Defense Survey.

$P_i=$ the estimated number of 17–20 year old physically and mentally qualified males (Mental Groups I-III) in the civilian labor force.[11]

$Y_i=$ relative earnings in the ith region—the ratio of (a) average annual military pay over the first term to (b) full-time civilian earnings of 16–21 year old males not enrolled in school in the ith region.[12]

$U_i=$ unemployment rate of 16–21 year old civilian males not enrolled in school in the ith region.[12]

Preliminary regressions indicated that pay elasticities were uniformly higher when enlistments were adjusted to eliminate the effects of the draft. A possible reason for the higher elasticity at lower levels of enlistments is that the services initially attract men who have strong preferences for military life. Those who remain in civilian life are less inclined toward the military and represent a smaller base from which the services may recruit. It therefore seems reasonable to expect that pay would have to be advanced by succeedingly higher increments to produce a given change in supply as enlistments move from low, no-draft to high, with-draft levels. This behavior is reflected in equation (2) where the pay elasticity falls as enlistment rates rise.[13] The following equations were estimated from 1963 regional data for Army and all-service enlistments:

---

portions as currently, even though the relative cost of labor would be higher. This assumption no doubt resulted in an overstatement of the deficits.

[11] This is a somewhat different population base than was used in the time series analysis. While there is little year-to-year fluctuation in the proportion of males who are physically and mentally qualified for military service or enrolled in school, there is considerable variation across regions. This must be accounted for in a regional analysis. To do so required a somewhat larger population base and thus the 18–19 year old group was expanded to 17–20 year old. See [4].

[12] Derived from U.S. Bureau of the Census Survey of Civilian Men, 16 to 34 years old, Oct., 1964.

[13] This equation is the complement of the equation in which enlistment rates are the de-

$$(\text{Army}) \log c_a = 4.7825 - .04297 \log Y^* - .01183 \log U \qquad (2a)$$
$$(.00940) \qquad (.00574)$$

$$\overline{R}^2 = .73$$

$$(\text{All Services}) \log c_t = 5.0334 - .10914 \log Y\dagger - .03540 \log U \qquad (2b)$$
$$(.03044) \qquad (.01859)$$

$$\overline{R}^2 = .64$$

NOTE: Figures in parenthesis indicate standard errors.
    * Significant at the .01 level.
    † Significant at the .05 level.

The negative signs for the pay and unemployment variables signify that regions with relatively high military to civilian incomes and high unemployment rates have higher than average enlistment rates. The relative pay elasticity is a function of the estimated regression coefficient of relative military pay and the enlistment rate. For example, in 1963 had there not been a draft the estimated Army enlistment rate[14] would have been 3.6 percent and the pay elasticity at that rate would have been 1.17; i.e., 10 percent change in military pay (civilian earnings held constant) would have increased Army enlistments by 11.7 percent. At the actual 1963 Army enlistment rate of 6.5 percent, the pay elasticity was equal to .62.

To derive estimates of the extent to which new officers could be attracted to military service through pay increases, we focused on variations in ROTC enrollment rates in voluntary institutions in 1963. The following equation was estimated from these data:

$$\log S_i = c_1 + c_2 \log Y_i + c_3 D_{A_i} + c_4 D_{N_i} + c_5 D_{AF_i} + c_6 D_{NS_i} \qquad (3)$$

where

S_i = "enrollment rate" in ROTC-I in institution i, 1963.

Y_i = 1963 earnings of male college graduates who graduated from institution i in 1958.

D_{A_i} = dummy variable representing the presence of an Army ROTC program in institution i.

D_{N_i} = dummy variable representing the presence of a Navy ROTC program in institution i.

D_{AF_i} = dummy variable representing the presence of an Air Force ROTC program in institution i.

D_{NS_i} = dummy variable representing institutions not located in the South.

---

pendent variable. The pay elasticity may be computed from the formula, $-b_2(1-R/R)$, where b_2 is the regression coefficient of the pay variable and R is the enlistment rate prior to the pay change. See [4].

[14] Mental Groups I–III Army enlistments as a proportion of the physically and mentally qualified 17–20 year old male full-time labor force.

In using this equation we assumed that the effects of expected military earnings and draft pressure were independent of the level of the enrollment rate and were reflected in the constant term. The dummy variables, D_{A_i}, D_{N_i}, D_{AF_i}, were used to control the effects of differing recruiting practices for ROTC among the services. D_{NS_i} was used to control for possible regional differences in tastes for the military. Data were used for only 82 of the 159 voluntary ROTC institutions because of inadequate sample size or because median earnings fell in open-ended class intervals.[15] The estimated coefficients are presented in equation (3a).

$$\log S = 8.869 - 1.987 \log Y^* + .2696 \, D_A^* - .2166 \, D_N^*$$

$$(.7940) \qquad (.0616) \qquad (.0571)$$

$$+ .1956 \, D_{AF}^* - .1585 \, D_{NS}^* \qquad R^2 = .41$$

$$(.0684) \qquad (.0737)$$

(3a)

NOTE: Figures in parentheses indicate standard errors.
    * Significant at the .01 level.

Earnings of college graduates were found to be negatively associated with enrollment rates.[16] Enrollment rates were also significantly related to the type of ROTC program and the geographic location of the institution.

On average, enrollment rates tend to be higher in schools offering Army or Air Force programs and in schools located in the South. The income elasticity was 1.99.[17] Unlike the enlisted analysis, this elasticity was assumed to be independent of the level of enrollment rates.[18]

Increased Payroll Costs Necessary to Attract an All-Volunteer Active Duty Force

In order to eliminate the recruitment deficits that would result if the draft were terminated in June, 1969, first-term pay would have to be increased by the amounts shown in Table 6. Since first-term military

---

[15] The earnings data were obtained from a survey of 1958 college graduates conducted for the NSF by the Bureau of Social Science Research in 1963. We are indebted to both of these organizations for their cooperation. Particular acknowledgement should be accorded to Mrs. Laure Sharpe and Mr. Rick Jones.

[16] It is possible, however, that the variable Y is picking the effects of an omitted variable, the proportion of graduates pursuing advanced studies. If these students are less likely to enroll in ROTC because of greater distastes for the military or a smaller degree of draft pressure, and they have higher incomes because of their advanced training, then the coefficient of Y will be biased upward. Unfortunately, an adequate measure of the proportion of graduates taking advanced degrees was not available for inclusion as an independent variable to control for its effects.

[17] Regressions estimated using linear and complement equations produced slightly higher elasticities for the no-draft world, but were poorer fits to the data [2].

[18] This assumption has, as a necessary condition, independence between draft effect and Y. Examination of survey data for this relationship did not uncover evidence to refute this assumption [2].

pay is most out of line with comparable civilian earnings for both enlisted men and officers and the implicit discount rate is relatively high for young men, the largest relative increases in military pay were allocated to military personnel within the first two to four years of service.

The estimates shown in Table 6 were based on the recruitment deficits shown in Table 5, and the pay elasticity estimates derived from equations (2a), (2b), and (3a).[19] The high and low estimates are one standard error above and below the estimated pay coefficients. The relatively large deviation between the "best" to the upper limit estimates results in part from the shape of the enlisted supply curve.

TABLE 6

ESTIMATED INCREASES IN FIRST-TERM PAY REQUIRED TO OBTAIN A
2.65 MILLION ALL-VOLUNTEER FORCE
(In Percent)

| | Enlisted* | | Officers† |
| | Unemployment Level | | |
	5.5%	4.0%	
Low.........	80	114	18
Best.........	111	163	26
High........	181	282	50

* Average increase in total tax equivalent income for all enlisted men during the first 1 to 4 years of service. Tax equivalent income includes all pay and allowances plus the tax advantage resulting from the fact that quarters and subsistence allowances are not subject to income tax.
† Average increase in officer tax equivalent income during the first 2 years of service.

Using 1964 military earnings as a base, increases in first-term pay for enlisted men averaged between $2,100 and $4,700 per year, at a 5.5 percent unemployment rate and between $2,900 and $7,300 at a 4.0 percent unemployment level. Since the Army would experience the largest recruitment deficit, an Army enlistment bonus was built into the estimates to minimize the amount of quasi-rent that might otherwise be paid to recruits in the other services and to career personnel. For first-term officers, pay increases would range between $900 and $2,500 per year. In addition to the pay increases provided to military personnel within the first two to four years of service, substantial increases in compensation for career personnel were provided in order to maintain a viable pay structure. These averaged between 15 percent and 90 percent for enlisted personnel and between 10 percent and 20 percent for officers.

In arriving at the total increase in payroll costs associated with attracting enough volunteers to maintain a 2.65 million force in the early

[19] We realize that cross-section data give a more accurate reflection of long-run supply responses to relative pay changes. However, we believe that these estimates are appropriate, since our primary concern was estimation of the long-run costs of eliminating the draft.

TABLE 7

ESTIMATED INCREASE IN PAYROLL COSTS NECESSARY TO OBTAIN AN ALL-
VOLUNTEER FORCE OF 2.65 MILLION*
(In Millions of Dollars)

	5.5% Unemployment			4.0% Unemployment		
	Low	Best	High	Low	Best	High
Enlisted:						
Increase in active duty pay.............	$3,570	$5,020	$8,960	$5,210	$7,640	$14,610
Increase in future retirement benefits* ..	160	390	1,030	420	790	1,940
Savings due to reduced turnover.....	−350	−410	−540	−410	−510	−720
Total..................	$3,380	$5,000	$9,450	$5,220	$7,920	$15,830
Officers:						
Increase in active duty pay.............	$ 270	$ 390	$ 760	$ 270	$ 390	$ 760
Increase in future retirement benefits† ..	40	50	100	40	50	100
Savings due to reduced turnover.....	−20	−20	−30	−20	−20	−30
Total..................	$ 290	$ 420	$ 830	$ 290	$ 420	$ 830
Total officers and enlisted...............	$3,670	$5,420	$10,280	$5,510	$8,340	$16,660

* Based on estimated recruitment deficits in the absence of the draft in fiscal year 1970–71.
† Assumes that the higher pay rates would not apply to those who had retired prior to the enactment of the new pay rates.

1970's, allowance was made for expected increases in retirement bene-
fits that would result from higher rates of basic pay and a larger propor-
tion of career personnel. Offsetting some of these cost increases, an all-
volunteer force would experience higher retention and thus lower turn-
over costs. Reduced turnover would permit the services to reduce the
number of men being trained each year and thus reduce the number
of military instructors and support-type personnel. By 1976, it is esti-
mated that military manpower could be reduced by 70,000 without
reducing the "effective" force level. The dollar savings would be small in
the first few years after termination of the draft. Thereafter, it would
grow to rather substantial sums and by 1976 would range between
$370 million and $750 million.

Our best estimates of the costs of recruiting an all-volunteer force in
the mid-1970's ranged from $5.4 to $8.3 billion, depending upon the
level of unemployment (Table 7). Comparison between the best and the
low- and high-cost estimates indicates that they are fairly sensitive to
changes in our estimates of pay elasticity, differing by from $1.8 to $8.3
billion when the pay elasticities were varied by one standard error. We

must reemphasize, however, that these figures apply only for active duty enlisted men and officer personnel exclusive of physicians and dentists, and they do not include any estimate of the costs of staffing the reserves entirely with volunteers.

The higher budgetary costs of an all-volunteer force reflect, in part, the income transfer that will occur between inductees and involuntary recruits, who are now bearing some of these additional costs implicitly, and the taxpayer, who will be required to bear these costs explicitly when we achieve an all-volunteer force. In addition, any manpower procurement system (which draws a large number of individuals) out of the civilian world involuntarily will have substantial effects on the civilian economy [5]. For example, our current draft system induces many potential draftees to seek ways of avoiding the draft, such as continued schooling, employment in draft-deferred occupations, or early marriage and fatherhood. Furthermore, many inductees and enlistees receive valuable training in the military which they might not be able to obtain in the civilian economy, thereby enhancing the skill mix of the civilian labor force. Finally, because the draft allows the military to employ first-term enlisted and officer personnel at less than competitive wage rates, it is likely that the military services overemploy labor, particularly high-quality labor [6]. We did not address these effects in our analysis, but we believe they are sufficiently important to warrant further study.

REFERENCES

1. W. Oi and S. Altman, "The Supply of Enlisted Personnel" (unpublished manuscript).
2. A. Fechter, "The Supply of New Military Officers," *Institute for Defense Analyses Study* (forthcoming).
3. *Review of the Administration and Operation of the Selective Service System*, Hearings before the Committee on Armed Services, House of Representatives, 89th Cong., 2nd sess., June, 1966.
4. S. Altman, "Earnings, Unemployment, and the Supply of Volunteers to Military Service" (unpublished manuscript).
5. W. Hansen and B. Weisbrod, "Economics of the Military Draft" (unpublished manuscript).
6. H. Gilman, "Military Manpower Utilization," *Defense Management*, edited by Stephen Enke (Prentice-Hall, Jan., 1967).

REENLISTMENTS IN THE U.S. NAVY: A COST EFFECTIVENESS STUDY*

By FRANKLIN M. FISHER, *Massachusetts Institute of Technology*
and ANTON S. MORTON, *Arthur D. Little, Inc.*

I. *Incentives and Reenlistment*

Navy men typically enlist for four years; then leave the Navy. Those who remain mostly stay until retirement. A low first-term reenlistment rate (FTRR) dissipates human capital—a training investment per man of six to twelve months.

The Navy felt it had a shortage of trained career enlisted men, especially in occupations which require extensive technical training, principally electronics maintenance and repair men. One obvious focal point of any attempt to alleviate such a shortage is the first-term reenlistment decision. Accordingly, our study began by estimating the several effects on the FTRR of a large number of possible incentives. These fell into six general classes: active duty pay; retirement pay; educational benefits; assignment, housing, and other policies; other fringe benefits; promotion opportunities. The present incentive system was also analyzed.

We estimated the effects of each incentive on the FTRR by survey data, calibrated to estimate actual FTRR's from statements of reenlistment intentions, assuming the institution of the incentives.

Except for the FTRR, conditional survival rates (the proportions of men in the Navy for t years who stay into their $t+$1st year) have been nearly constant for some time. We therefore assumed that the incentives would not affect these rates.[1]

II. *Costs and Benefits of Changing the Reenlistment Rate*

It is appropriate to choose that policy which will enable the Navy to perform its mission at least cost or that policy which will best enable the Navy to perform its mission at given cost. This may or may not involve the choice of that incentive which most raises the reenlistment rate, and it may not even involve raising the reenlistment rate at all.

[1] An exception was made in the case of incentives offering early retirement options. Detailed discussion of the estimation of incentive effects on reenlistment rates is given in Morton [3].

The effects of the adoption of a given incentive system may be conveniently separated into two parts: (1) direct costs associated with each man in the Navy under such a scheme; (2) costs and benefits involved in the effect the scheme has on the FTRR.

The direct costs are conceptually quite straightforward. Any incentive system involves expenditures over the career of each enlisted man; e.g., pay, fringe benefits, capital expenditures, and so forth. Their level depends on how long the man has been in the Navy. A different incentive system involves a different longevity-year pattern of direct costs.

Given the cost–longevity-year relationship, however, there is a further effect on costs which comes about through the altered FTRR. Raising it increases the probability that an enlistee will be in the Navy after t years (for t greater than 4 and less than some upper limit). Thus, the expected value of the costs to be incurred in his tth longevity year (costs times survival probability) is also increased. The expected present value of costs incurred on the induction of a new man is thus increased by an increase in the FTRR, even if costs for each longevity year do not change.

In return for the cost increase, the Navy acquires a greater proportion of experienced men. Since these are presumably more useful than inexperienced ones, the Navy's mission can be performed with fewer men, saving training and other costs of men who now need not be inducted.

Let us posit an indifference surface showing the combinations of experienced and inexperienced men which can perform a given task. By raising the FTRR, we substitute experienced for inexperienced men along this surface, with corresponding cost change.

While the true picture is more complicated, this concept produces an insight: since experienced men cost more than inexperienced ones, raising the FTRR need not reduce costs associated with the performance of a given mission. Even with a given direct cost structure, i.e., FTRR rises for exogenous reasons, a more expensive Navy performing a given mission may result. This actually happens with real data and parameters, as shown below.

III. *A Utility Function for the Navy: Marginal Rates of Substitution Among Experience Classes*

Since raising the FTRR substitutes fewer experienced for more inexperienced men, analysis must proceed by considering the marginal rates of substitution among men of differing experience, by trying to define a Navy utility function in terms of training levels.

Discussion with naval officers indicated that enlisted men could be grouped into four experience classes, with substitution within a class assumed on a man-for-man basis. The lowest experience class (the first)

consisted of all first-year men. Men after their first year were distributed among the remaining three classes in accordance with the actual distribution of men in each year-of-service in 1964 by groups of pay grades. We assumed that a typical group of inductees would advance in training to match the cross-section data for the base year.

The Navy's utility function was then defined with the sizes of the four classes as arguments. The usual consideration suggested diminishing marginal rates of substitution (e.g., rearrangement of personnel just compensating for the loss of an experienced supervisor becomes more difficult if such men are relatively scarce), and a four-factor Cobb-

TABLE 1
COBB-DOUGLAS EXPONENTS (β's)

Experience Class	All-Navy Assumption Set			Electronics Assumption Set	
	1	2	3	1	2
1. Basic.........	.268	.239	.255	.104	.086
2. Apprentice....	.386	.402	.393	.432	.444
3. Journeyman...	.268	.279	.273	.337	.341
4. Chief.........	.078	.081	.079	.126	.129

Douglas function (used ordinally only) was chosen for analytic convenience.[2] Thus, denoting the number of men in the ith experience class in calendar year t by Y_{it}, the Navy's utility function in year t was taken to be:

$$(1) \qquad U_t = Y_{1t}^{\beta_1} Y_{2t}^{\beta_2} Y_{3t}^{\beta_3} Y_{4t}^{\beta_4}; \qquad \sum_{i=1}^{4} \beta_i = 1.$$

To estimate the β_i, over 100 interviews were conducted with experienced supervisors scattered over nine ship types and six occupational groups, asking how many men of a given experience class would just compensate for the loss (or gain) of one man in another class; i.e., return them to the existing level of operational effectiveness. Given the men on-board, the implied β_i could then be determined for each interviewee from the marginal rates of substitution so described assuming his utility function to be of the form (1).

These results then had to be aggregated to obtain estimates for the overall Navy. This is not a simple matter. Experienced men in widely different jobs are not perfectly interchangeable; investment in human capital becomes embodied. There are two ways around this problem. First, because of the Navy's special interest in electronics personnel, we performed separate analyses for them and for the Navy as a whole.

[2] Our dynamic programming problem made such reasons a bit more compelling than usual. While only for this reason, a Cobb-Douglas was used, its parameters were varied considerably.

Second, we showed[3] that various assumptions about assignment of men to jobs justify estimating the β_i as weighted geometric averages of the interview-derived parameters, the weights being the number of men in each ship-occupational category. The analysis was performed both with the β_i so derived and with variations from them as given in Table 1.[4]

IV. *Incentive Cost: a Dynamic Programming Problem*

It would be unreasonable to require that we choose an incentive to maximize the utility function (1) subject to a cost constraint. (1) gives only the current utility function for year t. Since the number of men in each experience class is in part determined by past actions, the level of U_t at time t affects its level at later times. In principle, it might be possible to define a Navy utility function over the Y_{it} for all future times, but to do this in practice would be completely arbitrary. For example, to impose a discount rate and maximize the present value of a future utility stream imposes a cardinal form on the utility function. Up to this point, only various marginal rates of substitution have been observed. To maximize the present value of a future stream of U_t is to insist that (1) and not any monotonic transformation thereof is the proper representation of the Navy's indifference map. Lacking information on intertemporal marginal rates of substitution, this is distinctly unattractive. Fortunately, an ordinal alternative was available.

We set as our objective to minimize costs for given targets set in terms of U_t. Official Navy personnel requirements for the next thirty years were translated into values for the Y_{it} for those years. We then evaluated the U_t generated and required that any policy result in equal or greater U_t for each year. Subject to these constraints, we sought that incentive which minimized the present value of future costs.[5]

Since these costs depend upon induction rates[6] and since a wide variety of ways of meeting the given requirements exists, we wished *for each* incentive to find that induction pattern which minimizes present discounted costs. Those minimized values could then be compared for different incentives. Finding the optimal induction pattern can be formally described as follows:

[3] Details may be found in [4]. The human capital aggregation problem is closely related to the physical capital aggregation problem discussed in Fisher [1].

[4] While the standard errors of the estimated β_1 were very small (because of the large number of observations), the fit of utility functions using the estimated β_1 to the original interview data was very poor. This was largely due to a nonnegligible number of observations corresponding to man-to-man trade-offs scattered over the entire range of the observations. If these are ignored, as due to insufficiently thoughtful interview responses, the fit is much improved. If trade-offs really were man for man, our results below would be substantially strengthened.

[5] A discount rate of 5 percent was used. Previous experience suggested that results (in qualitative terms) are not sensitive to the choice of a discount rate.

[6] For simplicity, we assumed promotion and retirement rates fixed, so that, given the incentive package to be used, the only choice variables are induction rates.

Let I_t be the number of men inducted in year t under the given incentive program. Then:

$$Y_{1t} = I_t$$

(2)
$$Y_{it} = \sum_{\theta=1}^{t-1} \lambda_{it-\theta} I_\theta + B_{it} \qquad (i = 2, \cdots, 4)$$

Here the $\lambda_{it-\theta}$ are the fraction of the inductees of year θ in experience class i, in year t, while the B_{it} are the number of men in that experience class in that year who were already in the Navy at the start of the program (in year 0). These parameters are all derived from survival rates and the distribution of men in each longevity year over experience classes.

Let C be the present value of all program costs. Then C can be written as:

(3)
$$C = \alpha \sum_{t=1}^{30} \frac{I_t}{(1 + r)^t} + S$$

where: r is the discount rate; α is the expected present value of all perman costs associated with a man inducted at time 0 (costs in each future longevity year being weighted by the probability of survival into that year); and S is the similar expected present value of program costs assignable to men already in the Navy at the start of the program.[7]

For each incentive, we have the following nonlinear dynamic programming problem: Choose $I_1,\ldots, I_{30}$ so as to minimize (3) subject to (2) and

(4)
$$U_t \geq \overline{U}_t \qquad (t = 1, \cdots, 30)$$

where U_t is defined by (1) and $\overline{U}_t$ is the value of (1) in year t for a Navy just meeting stated personnel requirements in that year.

Since inducting each year exactly the number of men to meet requirements for that year was not necessarily the best policy (it may pay to stockpile men), this nonlinear programming problem had to be solved; accordingly, this considerable task was performed for each incentive under varying values of the parameters involved.[8]

V. Results

For the Navy as a whole, induction to meet current requirements turned out to be optimal in all cases. However, for every incentive such a myopic policy in electronics leads to a fantastically explosive and costly cycle in induction rates. The true optimal induction policy for electronics, on the other hand, involves a perfectly reasonable pattern of induction rates (at about historical levels) and a correspondingly

---

[7] Men inducted in year thirty, for example, are assigned the discounted value of all costs committed to them while they remain in the Navy. This avoids large end effects.

[8] Details of the solution method are given in [4] and summarized in [2].

appropriate level of costs (which suggests that the Navy is not myopic in this regard).

The difference is principally due to the high ratio of β_2 to β_1 in electronics and the large proportion of first-year electronics men in training.[9] With current inductions used to meet current targets exactly, a small drop (rise) in experienced men—a legacy from past decisions—must be compensated by a very large increase (fall) in new inductees.

The principal result is as follows. For every incentive, for both the Navy as a whole and for electronics, an increase in FTRR results in an increase in costs for a given effectiveness pattern. (Of course, substantially lowering the FTRR is inadvisable, because even with a draft, the Navy probably could not induct the required number of volunteers to fill the resulting annual gaps.) The Navy would do better to exercise selectivity in offering reenlistment incentives than to try to raise the the general reenlistment rate.

This result also shows up in the relative rankings of the incentives. The present incentive system ranks better than every other incentive except one, in electronics. This incentive is an option to retire after ten rather than twenty years of Naval service, and, of course, involves the assumption of a considerable reduction in the survival rate after ten years as well as an increase in the FTRR.[10]

Why should raising the reenlistment rate at no direct costs result in a more expensive Navy? A rise in the reenlistment rate means a substitution of fewer but more costly relatively experienced men for a greater number of cheaper inexperienced men. Such substitution, in the range of data and parameters studied, is cost increasing.

To illuminate the matter, suppose that inductions are is an optimal pattern and consider the induction of an additional 100 men at time t. This can be regarded as a production process which takes four years and results at time $t+4$ in some number (<100) career men. Denote costs involved in the production process by C_1.[11] There are also savings associated with the process, for the 100 men perform useful services in their first four years and thus enable the Navy to meet requirements in those years by inducting fewer men during those years. Call those savings S_1. The net costs of production of the career men involved is then $(C_1 - S_1)$, which may be positive or negative.

There are also costs and savings associated with the use of the elements of human capital so produced from $t+4$ on. As before, they are

[9] Students (and instructors) were removed from the Y_{it} by adjusting $\lambda_{it-\theta}$ for the proportion of the year spent in training.

[10] Considering only electronics personnel, two incentives offering rapid promotion to warrant office status may be a bit less expensive than the present incentive system (the caution is due to errors of approximation in the solution). Note that these too are incentives which keep men in the enlisted Navy after four years but encourage them to leave earlier than at present.

[11] Costs and savings in this discussion may be taken as discounted to time 0.

the direct costs associated with the career men and the savings which these career men exert on later inductions to meet later targets. Call those costs C_2 and the savings S_2; the net return from using the human capital is (S_2-C_2), which again may be positive or negative.

The total net return from the induction of the 100 men, therefore, is $(S_1-C_1)+(S_2-C_2)$. It is easy to see, however, that, if inductions are already optimal, this sum can never be positive, for, if it were, costs would be reduced by inducting the men. Indeed, the sum will be zero for years in which current requirements are more than fulfilled and will actually be negative for those years (of which there are quite a few, including year thirty) in which current requirements are a binding constraint. It thus cannot be true that (S_1-C_1) and (S_2-C_2) are both positive, and negativity tends to dominate.

Now consider a small increase in the reenlistment rate, other things equal. This produces in every year extra career men after a production process which takes four years but which costs nothing (for the production of the extra men). Thus, the net returns from doing this involve only (S_2-C_2). We know that these returns would fail to justify the production of the human capital at a cost of (C_1-S_1) which may be positive or negative; our result as to the effect of raising the reenlistment rate says that such production is not justified at zero cost, either.[12]

With existing later survival rates or pay scales, the Navy can accomplish a given mission more cheaply with a somewhat larger turnover of relatively inexperienced men than it can by encouraging those men to stay for an additional sixteen years. The fact that an early retirement incentive does reduce costs slightly in electronics suggests that the difficulty lies in a rate of pay and other benefits which rise with seniority in a manner not matched by rises in marginal products, so long as nearly all men staying after four years become career men who stay through twenty. This merits further investigation.

[12] It may be thought that choosing an optimal induction pattern for each incentive (on which choice this argument and, in part, the results depend) is an unrealistic process; it should be pointed out that these optimal patterns are not radically different from historical induction rates. This suggests that even in electronics (where, as opposed to the Navy as a whole, cost minimization is not merely a matter of myopic decision making), the Navy does make some attempt to optimize. Even aside from this, it is reasonable to require a rational policy choice to involve optimization of this sort. If the Navy is far from an optimal induction pattern, changing that is likely to have more effect on costs than raising the reenlistment rate.

REFERENCES

1. F. M. Fisher, "Embodied Technical Change and the Existence of an Aggregate Capital Stock," *Rev. of Econ. Studies*, Oct., 1965, pp. 263–88.
2. F. M. Fisher and A. S. Morton, "The Costs and Effectiveness of Reenlistment Incentives in the Navy," *Operations Research* (forthcoming).
3. A. S. Morton, "Predicting the Impact of Policy Changes from Surveys" (paper read at the American Meeting of the Institute of Management Sciences, San Francisco, Feb., 1965).
4. A. S. Morton, F. M. Fisher, and D. M. Nitzberg, *Cost/Effectiveness of Reenlistment Incentives* (Inst. of Naval Studies, Cambridge, Mass., Jan., 1966, for official use only.)

THE ECONOMIC COST OF THE DRAFT*

By WALTER Y. OI

University of Washington

The draft represents one means of supplying the armed forces with qualified personnel. Under the current draft, accessions to military service are of three types: (1) true volunteers who freely choose military service over alternative civilian job opportunities, (2) reluctant volunteers who enlist in preference to being drafted, and (3) draftees who are involuntarily inducted. The last two groups are coerced to serve by the military service obligation that is imposed by the current draft law.

In times of war when nearly everyone must serve to assure the defense of the nation, alternatives to a draft are judged to be too costly or infeasible. Peacetime demands for military personnel are, however, considerably smaller, with the consequence that a draft becomes selective. Debates over the equity of the selection process (which under the current draft translates into deferment policies) are symptomatic of a search for an alternative to current military manpower procurement policies. To say that a particular alternative such as an all-volunteer force[1] is preferable to the current system implies that the cost of the alternative is, in some sense, lower than the cost of the current draft. If the draft were abolished, military pay and other recruitment incentives must be improved to attract sufficient recruits to meet prescribed military manpower objectives. In his statement before the House Armed Services subcommittee, the Hon. T. D. Morris (Assistant Secretary of Defense) indicated that an all-volunteer force of 2.65 million men would increase the annual military payroll budget by $4 to $17

[1] The current draft law (the Universal Military Training and Service Act of 1951 as amended and extended) expires on June 30, 1967. The proposed alternatives to an extension of the present law include (1) lottery at a younger age of induction, thereby shortening the period of draft liability, (2) universal military service, (3) equivalent national service wherein some youths could serve in the Peace Corps, VISTA, or other government sponsored programs, and (4) establishment of a voluntary professional army.

billions.[2] The military payroll of the Department of Defense (hereafter abbreviated DOD) is not the economic cost of labor resources allocated to the uniformed services. The draft has surely affected both the level and structure of military pay. Moreover, many of the men who serve are conscripted or are recruited under the threat of a draft liability. In addition to the costs borne by those who do serve, it is argued that the uncertainty of being drafted creates other real and psychic costs for those who avoid military service by obtaining deferments.

The cost of acquiring and retaining military personnel can be measured in several ways of which the budgetary cost is one. The financial cost to the economy is defined in this paper as the value of civilian outputs that could have been produced by the labor resources which were allocated to the armed forces. This concept which completely ignores occupational preferences provides a measure of technical efficiency in terms of civilian outputs that were foregone to achieve given levels of military preparedness. The full economic cost of the draft must, however, acknowledge occupational preferences for military versus civilian employments. If an individual has an aversion to service life, he could, in principle, be compensated by enough to induce him to volunteer. Presently, individuals who would require such compensation are forced to serve by the draft law. In this paper, I propose to compare these costs for two hypothetical forces with the same active duty strength of 2.65 million men: one a purely voluntary force and the other a mixed force composed of conscripts, true and reluctant volunteers.

I. *Force Strengths and Military Manpower Requirements*

The labor resources demanded by the armed forces can be measured by force strengths. The force strength is simply a stock demand for military personnel unadjusted for the quality of servicemen or the proportion in an effective status.[3] The total defense establishment can conveniently be divided into three forces: (1) officers on active duty, (2) enlisted men on active duty, and (3) paid drill reserves. The last component engages in active duty for training only and is rarely used to bolster active duty strengths.[4]

In the six years prior to the Vietnam build-up, the average strength

---

[2] House of Representatives, 89th Cong., Second Sess., "Review of the Administration and Operation of the Selective Service System," Hearings before the Committee on Armed Services (June 22, 23, 24, 28, 29, and 30, 1966) (hereafter abbreviated as *House Hearings*), pp. 9923–59; see especially pp. 9936-40.

[3] The effective strength is defined as the number of men who are not in a "training status." Men who are being trained or who are engaged in training others are deducted from total force strength to arrive at the effective strength. A lower personnel turnover would therefore lead to a larger proportion in an "effective" status.

[4] In the last decade, reservists were recalled to active duty in significant numbers only once. During the Berlin crisis of FY 1962, some 111 thousand Army and National Guard reservists were activated to raise Army force strengths from 858.6 to 1,066.4 thousand.

in all active and reserve components was just under 3.7 million men of which 2.6 million were on active duty (see Table 1). The fluctuations in active duty strengths are largely due to international tensions. Since a discussion of the factors which determine force strength objectives is beyond the scope of this paper, the peacetime strength objectives are taken to be exogenous.

A more meaningful demand concept is provided by the gross flow demand for accessions from civilian life. It is the number of required accessions A_t that must be recruited or conscripted to replace losses during the year L_t and to achieve changes in stock demands ($S_t - S_{t-1}$).

$$A_t = L_t + (S_t - S_{t-1})$$

If force strength S_t is held constant, required accessions must equal losses which arise because of voluntary separations upon completion of obligated tours of duty, retirements, deaths, and discharges for medical and unsuitability reasons.

The military manpower procurement channels which have evolved under a draft have strongly influenced the characteristics of servicemen. Of the 645 thousand annual accessions to military service in FY 1960-65, 539 thousand (83.6 percent) entered through a variety of voluntary programs. An individual can discharge his service obligation by entering active duty forces as an enlisted man or officer. He can accomplish the former as a voluntary enlistment or as an involuntary draftee, usually to the Army.[5] Except for the doctors and dentists draft, all officer procurement programs are voluntary. Finally, the draft liability can be satisfied by entering a reserve or National Guard unit which requires active duty only for training. The historical trends in accessions and projected accessions in FY 1970-75 under a continued draft are shown in the lower panel of Table 1. In order to extrapolate the characteristics of men who are likely to enter military service in the future, it is convenient to study the service experience of age classes (cohorts) born in specific years.

The disposition of military service obligations by men born in 1938 was estimated from a sample of Selective Service registrants.[6] By July, 1964, 51.6 percent of this age class had some active or reserve military service with the highest participation rate of 59.5 percent for men with some college education. Given current physical, moral, and mental

---

[5] A few delinquents (mainly reservists who fail to attend drill meetings) are drafted into the other three services. The Navy drafted some men in FY 1956, and the Marines in FY 1966. Some voluntary enlistments serve in two-year active duty reserve programs offered by the Navy and Marine Corps. The Army offered a two-year reserve enlistment in 1956, but the program was dropped in 1958.

[6] Confer *House Hearings*, p. 10011.

TABLE 1

FORCE STRENGTHS AND ACCESSIONS FROM CIVIL LIFE

(Actual FY 1960–65 and Projected FY 1970–75, in Thousands)

	1960	1961	1962	1963	1964	1965	Annual Averages	
							1960–65	1970–75*
Force strengths†								
DOD (active duty) total	2,476.4	2,483.8	2,807.8	2,697.7	2,685.2	2,653.1	2,634.0	2,650.0
Officers	316.7	314.8	343.1	333.4	336.4	337.6	330.3	340.0
Enlisted	2,159.7	2,168.9	2,464.7	2,364.3	2,348.8	2,315.5	2,303.7	2,310.0
Army (active duty) total	873.1	858.6	1,066.4	975.2	972.4	968.3	952.3	969.5
Officers	101.2	99.9	116.1	107.8	110.3	111.5	107.8	112.5
Enlisted	771.8	758.7	950.4	867.4	862.2	856.8	844.6	857.0
Reserves and National Guard (paid drill)	1,079	1,086	958	964	1,048	1,006	1,023.5	—
All components total	3,555.4	3,569.8	3,765.8	3,661.7	3,733.2	3,659.1	3,657.5	—
Accessions from civil life								
DOD total	469.8	475.3	622.4	488.0	569.3	495.0	520.0	507.7
First enlistments‡	349	386	423	373	377	352	376.7	416.7
Inductions‡	90	60	158	74	151	103	106.0	55.3
Officers§	30.8	29.3	41.4	41.0	41.3	40.0	37.3	35.7
Army	206.3	188.6	303.2	203.0	285.8	221.0	234.5	228.5
First enlistments‡	106	118	127	113	117	103	114.0	159.0
Inductions‡	90	60	158	74	151	102	105.8	55.3
Officers§	10.3	10.6	18.2	16.0	17.8	16.0	14.8	14.2
Reserves and National Guard#	130.0	130.0	90.0	110.0	170.0	120.0	125.0	—
Total entries all components	599.8	605.3	712.4	598.0	739.3	615.0	645.0	—

* SOURCE: *House Hearings*, p. 9954.
† Active duty force strengths include both sexes as reported in *Statistical Abstract of the United States 1966*, Table 365, p. 261 (Government Printing Office: Washington, D. C., 1966). The paid drill reserve and National Guard strengths were obtained from unpublished data, DOD Statistical Office.
‡ First enlistments include two-year reserve enlistments but exclude reserves recalled to active duty. Confer *Statistical Abstract, ibid.*, Table 366, p. 262.
§ Data for officer accessions taken from a special tabulation prepared by the DOD Statistical Office.
# *House Hearings*, p. 10001.

qualification standards for military service, roughly 30.4 percent of these men would have been rejected.[7] From independent data, I estimated that 59.2 percent of qualified males entered active duty forces while another 14.4 percent served in reserve and Guard units. The incidence of active military service also varied across education groups from a high of 76.8 percent of qualified high school graduates to 32.3 percent of qualified college graduates.

The projected flows of accessions in FY 1970-75 under a continued draft were developed in the DOD study and appear in the last column of Table 1.[8] These projected flows were juxtaposed to a typical age class born in 1946-48 which is estimated to contain 1,880 thousand males. The estimated participation rate in active military service falls from 41.5 percent of men born in 1938 to 27.0 percent for the age class of 1947. Only 38.5 percent of qualified males will be needed to staff active duty forces in FY 1970-75.

Before turning to the characteristics of accessions to a voluntary force, it is desirable to assemble data on the structure of the entire mixed force. Higher qualification standards and secular gains in the educational attainment of the entire population have contributed to an upward trend in the average education of members of the armed forces. The percentage of enlisted men with high school degrees climbed from 55.2 to 72.8 percent between 1956 and 1963. Over the same period, the percentage of officers with college degrees rose from 55.5 to 69.4 percent. From the 1960 Population Census, it was possible to obtain the age and educational distribution of members of the armed forces stationed in the United States. If these data are expanded to the assumed force strength of 2.65 million men, I obtain the distribution shown in Table 2.[9] In relation to the male civilian labor force, the armed services have smaller fractions in the lowest educational level due to the mental qualification standards. The proportion

[7] Approximately 16 percent of an age class is found to be physically unfit, while an additional 2.5 percent are rejected for moral reasons, mainly habitual criminals. These standards have remained quite stable over the last twenty years. Mental qualification standards have, however, been raised as supplies of enlistment applicants grew in relation to requirements. Currently, mental standards have been lowered to a percentile score of 16 on the Armed Forces Qualification Test. The overall rejection rate for all reasons was 40.8 percent for men with 0-11 years of education and 19.1 percent for college graduates.

[8] These projections assume an unemployment rate of 5.5 percent, the average unemployment rate between 1957 and 1964. For details of the DOD projections, see *House Hearings*, p. 9954. If the unemployment rate is as low as 4.0 percent, voluntary enlistments can be expected to fall with an accompanying rise in inductions.

[9] The total DOD force strength as of June 30, 1960, was 2,447 thousand males, but the Census enumerated only 1,715 thousand. The distribution shown in Table 2 thus invokes two assumptions. First, it is assumed that the same age-education distribution applies to men stationed in the U.S. and overseas. Second, if the draft is extended, future accessions and reenlistments will generate the distribution which was observed in 1960. I believe that these two assumptions impart a downward bias to the educational attainment of mixed forces in FY 1970-75.

TABLE 2

DISTRIBUTION OF ARMED FORCES BY AGE AND EDUCATION
(For Force Strength of 2,650; Figures in Thousands)

Age	Years of School Completed					
	0–8	9–11	12	13–15	16+	Total
Total DOD						
17 or less......	11.3	55.2	6.7	.4	—	73.7
18–19.......	21.7	141.3	218.6	14.3	.2	396.2
20–21.......	18.4	111.3	235.1	48.3	2.7	415.8
22–24.......	30.8	97.9	238.3	81.1	79.6	527.7
25–29.......	34.7	89.3	181.0	47.4	63.1	415.5
30–34.......	38.2	67.8	124.3	34.7	28.5	293.5
35–44.......	44.2	67.2	178.2	81.2	59.9	430.8
45–54.......	12.3	12.0	21.2	14.9	21.3	81.8
55 and over ...	2.8	3.2	4.3	1.6	3.2	15.1
Total	214.5	645.2	1,207.8	323.9	258.5	2,650.0

SOURCE: Derived from *U.S. Census of Population, 1960*; Subject Report: "Educational Attainment," Code No. DC(2) 5B, Table 4, p. 54.

of college graduates is also somewhat lower than in the civilian sector for the age groups above 30. Finally, since military pay is primarily a function of years of service, the age structure (in terms of years of service) of the force as of June 30, 1965, is presented in Table 3. Under a continued draft, it is probable that the retention profiles of enlisted and officer personnel will not alter so that the mixed force of FY 1970-75 can be expected to exhibit a similar age structure.

The task of estimating the cost and composition of an all-volunteer force was approached in two steps. It was first assumed that the draft would be abolished with no accompanying changes in pay or other recruitment incentives. An implication of this exercise is that supplies of

TABLE 3

ACTIVE DUTY FORCE STRENGTH BY YEARS OF SERVICE
(As of June 30, 1965, in Thousands)

Years of Service	Total	Officers	Enlisted
0..........	427.7	28.4	399.3
1..........	488.4	34.1	454.3
2..........	301.6	25.0	276.5
3..........	226.3	20.1	206.2
4..........	99.8	12.9	86.8
5–9.........	321.6	54.0	267.7
10–14........	338.6	51.1	287.4
15–19........	291.0	58.4	232.6
20–24........	131.4	45.4	86.0
25 and over	17.7	9.4	8.1
Total........	2,643.8	338.8	2,304.9

volunteers fall short of requirements for the strength objective of 2.65 million men. In the second step, military pay was advanced to attract enough volunteers to meet the strength objective.

If the draft is eliminated, the services would obviously lose the draftees who had accounted for 21 percent of accessions to enlisted ranks in FY 1960-65. In addition, the reluctant volunteers who had enlisted because of the draft liability are likely to remain in civilian life. To determine the probable number of reluctant volunteers, the Department of Defense made a survey in the fall of 1964 of servicemen and civilians in the military ages of 16-34. First-term regular enlisted men (who had all voluntarily entered service between 1960 and 1964) were divided into true and reluctant volunteers on the basis of their responses to the question: "If there had been no draft, and if you had no military obligation, do you think you would have volunteered for active military service?" Those who replied "no definitely" or "no probably" were classified as reluctant. The percentage of true volunteers in subgroups identified by age and education at time of enlistment are shown in the first column of Table 4. The proportion of true volunteers is highest in the youngest and least educated group and declines with age and educational attainment. The fourth column gives the estimated annual flows of voluntary enlistments in FY 1970-75 if the draft is continued, the mixed force case. Multiplying by the proportions of true volunteers, I obtain the estimated flows of enlistments if the draft were abolished with no pay changes.[10] The same question on draft-motivation was asked of officers on their first obligated tours and revealed that 41.3 percent of officer accessions were reluctant volunteers. In the absence of a draft, it is probable that officer procurement programs will be revised placing more reliance on noncollege graduate sources. Space precludes a fuller discussion of the supplies of officer personnel.

Depletions in supplies of initial accessions due to the loss of draftees and reluctant volunteers are offset to some degree by improvements in retention. Air Force studies show that first-term reenlistment rates are substantially higher for airmen whose primary reason for entry was unrelated to the draft.[11] If the reenlistment rate of reluctant volunteers is assumed to be the same as that of draftees, one can deduce probable reenlistment rates in an all-volunteer force.[12] Higher reenlist-

---

[10] These estimates contain a downward bias. It is likely that by FY 1970-75, the other services will enjoy excess supplies of enlistment applicants, some of whom will replace the reluctant volunteers. However, the shortfall in regular Army enlistments will persist.
[11] Survey questionnaires try to determine the single most important reason for original entry. Two of several possible choices include "choice of service" and "volunteered in preference to being drafted." Follow-up studies of respondents reveal that airmen selecting these responses had substantially lower reenlistment rates.
[12] Over the period 1957-64, the weighted average for first-term reenlistment rates of

ment rates imply lower personnel turnover which, in turn, leads to smaller required accessions to sustain a given force strength. If the

TABLE 4

EFFECT OF THE DRAFT ON VOLUNTARY ENLISTMENTS—SURVEY RESPONSES
(Classified by Age and Education)

Age at Entry and Education	Percentage of True Volunteers*	Number of DOD Sample†		Voluntary Enlistments in FY 1970–75		
				With Draft‡	No Draft	
		Number	Percent		Number§	Percent
17–19 years of age						
Less than high school.	79.3	167.8	27.7	122.2	96.2	36.6
High school graduate.	63.7	247.1	40.8	188.0	119.7	45.5
Some college.........	55.9	44.0	7.3	18.3	10.2	3.9
Total..............	68.7	458.9	75.8	328.5	226.1	86.0
20 and over						
Less than high school.	60.2	20.2	3.3	14.3	8.6	3.3
High school graduate.	42.3	61.7	10.2	42.8	18.1	6.9
Some college.........	32.7	64.4	10.6 ·	31.1	10.2	3.9
Total..............	40.5	146.4	24.2	88.2	36.9	14.0
All ages						
Less than high school.	77.4	31.1	31.1	136.5	104.8	39.8
High school graduate.	59.5	51.0	51.0	230.8	137.8	52.4
Some college.........	42.1	17.9	17.9	49.4	20.4	7.8
Total..............	61.9	100.0	100.0	416.7	263.0	100.0

* Based on responses of regular enlisted men in their first term of service to the question, "If there had been no draft and if you had no military obligation, do you think you would have volunteered for active military service?" Entries denote the percentage who responded, "Yes, definitely," or, "Yes, probably."
† Figures may differ from force strength statistics due to elimination of nonrespondents and sampling variability.
‡ Estimates of voluntary enlistments in FY 1970–75 if the draft is continued.
§ Obtained by multiplying columns 1 and 4. Assumes that the draft is eliminated but pay and recruitment incentives are unchanged.

transitional problems of moving to a voluntary force are ignored, I obtain the following estimates of gross flow demands (required accessions) to maintain a force of 2.65 million men:

draftees was $R_d = 7.67$ percent as compared to $R = 22.06$ percent for regular Army enlistees. The observed rate for all enlistees, R, can be regarded as a weighted average of the rate for true volunteers R_v and of reluctant volunteers R_d which is assumed to be equal to that of draftees.

$$R = kR_v + (1 - k)R_d$$

where k is the proportion of true volunteers. In the case of the regular Army where $k = .568$ (the proportion of true volunteers), the first-term reenlistment rate is estimated to rise from $R = 22.1$ percent to $R_v = 33.0$ percent. Similar improvements in retention can also be expected of the other services as well as for officers.

GROSS FLOW DEMANDS FOR A FORCE STRENGTH OF 2.65 MILLION MEN

Component	Voluntary Force (no draft)	Mixed Force (with draft)	Ratio
Total required accessions to enlisted ranks..	333.5	472.0	.707
Other services.........................	188.9	257.7	.733
Army—total.........................	144.6	214.3	.675
Volunteers*.........................	144.6	159.0	.909
Inductions.........................	0	55.3	—
Required accession of officers	28.4	35.7	.796
Total gross flow demands.................	361.9	507.7	.713

* Assumes a 5.5 percent unemployment rate.

Voluntary supplies in the absence of a draft fall short of requirements for the strength objective with the deficit being largest in the Army.

The manpower deficits can be eliminated by raising military pay to attract more volunteers. The responsiveness of enlistments to pay changes was estimated from cross-sectional data.[13] Of several functional forms which were fitted to the data, the most consistent fit was provided by a complement supply equation.[14]

$$(1) \qquad\qquad 1 - E = \alpha Y^{-\beta}$$

where E is the enlistment rate and Y denotes relative pay, the ratio of first-term military pay M to alternative civilian pay C. This supply equation implies that a 1 percent rise in relative pay leads to a β percent decline in the remainder (complement) of the population not in military service. The elasticity of supply ε is,

$$(2) \qquad\qquad \epsilon = \beta \left(\frac{1 - E}{E} \right)$$

An upper bound to the necessary pay increase is obtained by restricting the analysis to Army enlisted personnel. For a total strength of 2.65 million, the Army enlisted strength is 857 thousand. Under a continued draft, 159 thousand recruits are projected to volunteer while

[13] True voluntary enlistment rates in relation to the civilian male labor force 17-20 years of age (adjusted for mental qualification) were estimated for the nine Census regions from the 1964 DOD survey. Two enlistment rates were developed corresponding to (1) total DOD enlistments in Mental Groups I to III and (2) Army enlistments in Mental Groups I to III. Voluntary enlistments of men in upper Mental Group IV (AFQT percentile scores of 20-30) were omitted since their enlistments were limited by recruitment quotas. The data which were used in the supply analysis can be found in *House Hearings*, p. 9957.

[14] This functional relationship was suggested to me by Prof. A. C. Harberger. In estimating the parameters, an unemployment variable was included as an explanatory variable. I have deleted it here to simplify the exposition. Since unemployment is held constant in the projections, its influence is included in the constant term α. A fuller analysis of supplies of military personnel is contained in the paper by Altman and Fechter in this series.

another 55.3 thousand will be drafted. In a steady state after the transition, a voluntary Army with its lower personnel turnover can be sustained by annual enlistments of only 144.6 thousand. If the draft is eliminated, Army enlistments are expected to fall by 43.2 percent to annual inputs of 90.3 thousand. The necessary increase in enlistments is thus given by the ratio of required accessions to voluntary supplies with no pay changes; that is, 144.6/90.3 $=$ 1.601. The pay increase (Y_1/Y_0) that will achieve this increase in enlistments is estimated from the complement supply equation.

$$(3) \qquad \left(\frac{1-A}{1-E}\right) = \left(\frac{Y_1}{Y_0}\right)^{-\beta}$$

The estimated Army supply equation revealed an estimate for β of .108 implying an initial elasticity of supply of 1.36. The necessary pay increase to meet manpower requirements on a voluntary basis was estimated to be 1.68.[15]

The 68 percent pay increase is presumed to apply to men on their first term. The mechanics of implementing this pay increase deserve brief mention. The annual income of a serviceman includes money payments for various reasons and some income in kind. In this paper, I shall use the concept of tax equivalent income which includes (1) base pay, (2) money payments for subsistence, quarters, and uniforms, (3) the implicit value of subsistence and quarters if the serviceman receives no money allowances for these, and (4) the tax advantage.[16] The total military incomes of enlisted men classified by years of service appear in Table 5. It will be noticed that pay is extremely low in the first two years because pay increases prior to 1964 applied mainly to the career force. If first-term pay is raised by 68 percent, I assumed that the discontinuity in the pay structure would be eliminated so that the average pay over the first three years of service would climb from $2,500 to $4,200. In order to prevent reversals in the pay structure, the career force would enjoy a 17 percent increase in annual tax equivalent incomes.

With these pay increases, the armed services should be able to meet

---

[15] My estimate of the necessary increase in first-term pay is lower than the DOD estimates. The DOD study presented three estimates corresponding to three values of β; namely, the point estimate of β and the point estimate plus or minus one standard error of the regression coefficient. For the case of a 5.5 percent unemployment rate, I have reproduced the estimated percentage increases in first-term pay; see *House Hearings*, p. 9958; DOD low estimate, 80; DOD best estimate, 111; DOD high estimate, 181; my estimate, 68. The difference between the DOD best and my estimates is due to different estimates of required accessions. My procedure deals with a steady state in which the Army enjoys the retention profile of a truly voluntary force.

[16] Income in kind is clearly nontaxable. In addition, subsistence, quarters, and uniform allowances are classified as nontaxable, thereby giving servicemen an indirect tax advantage which accounts for about 5.5 percent of tax equivalent income.

TABLE 5

ANNUAL MILITARY INCOMES OF ENLISTED MEN
(For Pay Scales of FY 1963)

Years of Service	Total Income DOD	Army			Base Pay as Percent of Total Income
		Total Income	Taxable Income	Base Pay	
1......	1,830	1,900	1,058	1,055	55.5
2......	2,143	2,304	1,359	1,382	60.0
3......	2,991	3,247	2,199	2,002	61.7
4......	3,344	3,711	2,392	2,433	65.6
5......	4,130	4,248	2,691	2,575	60.6
6......	4,462	4,465	2,792	2,725	61.0
7......	4,649	4,596	2,937	2,858	62.2
8......	4,741	4,797	3,037	3,003	62.6
9–12....	5,235	5,377	3,409	3,280	61.0
13–16....	5,926	6,043	3,918	3,885	64.3
17–20....	6,387	6,414	4,245	n.a.	—

strength objectives with voluntary supplies of personnel. Since the incidence of reluctant volunteers is higher for older, more educated men, the voluntary force can be expected to have a lower average educational attainment. Because of its lower personnel turnover, only 27.5 percent of qualified males is required to sustain a voluntary force of 2.65 million men.

The lower personnel turnover also raises the career ratio defined as the proportion of men on their second and later tours of duty. Hence, the move to a voluntary army raises the career ratio of regular Army enlisted men from .431 to .537. The age structure of the voluntary force was thus developed from the new retention profiles and appears in Table 6. Finally, the educational distribution of the voluntary force was inferred from the education mix of true volunteers. It was assumed that officer procurement programs for noncollege graduates would be expanded. The proportion of enlisted men with 0-8 years of education was small for two reasons. First, the minimum AFQT score is assumed to be held at 16, and second, the secular trend toward more education implies a slower growth for this population base. The hypothetical educational distribution of Table 7 pertains to a voluntary force in a steady state after the transition period.

In analyzing the cost and characteristics of a voluntary force, I neglected several pertinent factors of which at least three should be discussed: (1) the transitional period, (2) the savings which obtain from lower turnover, and (3) possible substitutions of civilians for uniformed men. In order to maintain force strengths in the transition, required accessions are likely to be at least 10 percent greater than in the steady state. The necessary first-term pay increase to sustain Army

TABLE 6

ACTIVE DUTY FORCE STRENGTH BY YEARS OF SERVICE
(Hypothetical Voluntary Force, in Thousands)

Years of Service	Total	Officers	Enlisted
0............	355.3	34.5	320.8
1............	327.5	33.6	293.9
2............	299.2	32.1	267.1
3............	204.4	20.2	184.2
4............	108.0	16.0	92.0
5–9..........	455.6	65.4	390.2
10–14.........	343.3	51.5	291.8
15–19.........	304.6	44.1	260.5
20–24.........	165.2	28.1	137.1
25 and over	86.9	14.5	72.4
Total.........	2,650.0	340.0	2,310.0

force strengths during the transition is estimated to be 94 percent as opposed to the previous 68 percent pay increase. The average annual military pay over the first term climbs from $4,200 to $4,850.

In the mixed force, at least 10 percent of the active duty force is involved in training. Since initial accessions to a voluntary force are some 30 percent smaller, the voluntary force strength could be reduced by at least 3 percent and still retain the same number of men in an "effective" status. In addition, if men who are engaged in training others are moved to other duty assignments, further savings could be realized. A cut in the size of a voluntary force which achieves the same effective strength implies a corresponding reduction in required accessions. Neglect of these savings from lower turnover thus imparts an upward bias to my estimate of the necessary pay increase.

Many medical, clerical, food service, and maintenance positions which are now staffed by servicemen could be filled by civilians. Given

TABLE 7

ACTIVE DUTY FORCE STRENGTH BY AGE AND EDUCATION
(Hypothetical All-Volunteer Force)

Age	Years of School Completed					
	0–8	9–11	12	13–15	16 and Over	Total
Total DOD						
17–19 ...	25.7	145.1	212.1	18.0	—	400.9
20–21 ...	30.3	170.8	248.6	26.5	4.8	481.0
22–24 ...	28.8	133.9	194.0	31.8	41.1	424.6
25–29 ...	22.6	127.3	188.7	41.0	84.6	464.2
30–34 ...	17.1	100.5	147.5	29.7	52.5	347.3
35–44 ...	20.7	124.1	182.7	39.6	74.5	386.9
45–54 ...	3.1	19.1	28.8	8.6	30.2	89.7
55 and over.	—	—	0	0	0.7	0.7
All ages...	143.3	820.8	1,202.4	195.2	288.3	2,650.0

current military pay scales, assignment of military personnel to these jobs may well produce the lowest budgetary cost. However, when military pay is sharply advanced, it becomes economical even from the viewpoint of budgetary cost to substitute civilians for uniformed men.[17] By implementing these substitutions, the size of the armed forces could be reduced without changing the tasks performed by the totality of civilians and military personnel. A careful study of these possibilities for substituting civilians should, in my opinion, be an integral part of moving toward a voluntary force.

II. *The Financial Cost of the Armed Forces*

According to the defense budget for FY 1965, the cost of active duty military personnel was $12,662 million; retirement benefits accounted for another $1,384 million.[18] A second estimate of the payroll cost is obtained by applying the annual military incomes of Table 5 to the age structure of the force. This estimate of $12,049 million differs from the DOD cost because my measure of military pay excludes many pay items.[19] This latter procedure is, however, the only way to estimate the budgetary cost of a voluntary force.

In order to achieve a voluntary force with the same strength as the mixed force, the entire pay profile is shifted upward with the largest pay increases applying to the first four years of service. In a voluntary force, larger fractions of men are in the career force with the result that the higher military incomes in later years receive more weight. If the higher pay profiles are applied to the probable age structure of a voluntary force, Table 6, I obtain an estimated payroll cost of $16,103 million.[20] These cost estimates thus imply that the defense

[17] The stock demand for military personnel is not completely inelastic. At higher prices, fewer men would be demanded by the services. A consideration which operates in the opposing direction is the provision of rotational billets. Some jobs must be kept for men on rotation from overseas assignments. The demand for military men is likely to be determined by the anticipated peak load demand so that a considerable part of the force is always idle. This is as it should be and the reason for maintaining a standing army.

[18] *Statistical Abstract of the U.S.*, 1966, Table 153, p. 112.

[19] Flight pay, sea duty pay, and many other pay items were omitted in the annual tax equivalent incomes shown in Table 5. Another source of error is that the age structure of the force, Table 3, applies to years of active military service, while the pay profile of Table 5 is based on years of service for pay purposes; the latter is always equal to or greater than the former, thereby imparting a downward bias to my estimate of the payroll cost. Finally, I had only rough estimates of annual tax equivalent incomes for officers.

[20] In arriving at this cost, I assumed that an enlisted man received $3,900 in his first year of service with annual increases of $300 over the next three years. Enlisted men's pay in later years was raised by 17 percent. The officers' pay profile was adjusted to eliminate the discontinuity in the pay profile. As a result, average annual pay for officers in their first three years of service rose by 20.1 percent. The additional payroll cost of a voluntary force contains an upward bias since I have neglected the savings from lower personnel turnover and possible substitutions of civilians for uniformed personnel. Finally, it may be cheaper to attract recruits with recruitment incentives other than higher annual military incomes. Initial enlistment bonuses, greater pay differentials for skilled personnel, or educational/training benefits could be incorporated into the pay structure.

TABLE 8

MEDIAN INCOMES OF CIVILIAN MALE LABOR FORCE, 1964

Age	Years of Education Completed				
	0–8	9–11	12	13–15	16+
17–19....	2,010	2,926	3,196	3,147	—
20–21*...	2,391	3,314	3,924	4,668	—
22–24†...	3,160	4,026	4,789	5,168	5,280
25–29‡...	3,673	4,500	5,366	5,502	6,213
30–34....	4,296	5,339	6,167	6,910	8,353
35–44....	4,710	5,860	6,528	7,389	9,853
45–54....	4,717	5,636	6,549	7,855	10,846
55 and over .	4,229	4,944	6,135	6,642	9,883

* Incomes for males 21 years of age and under estimated from DOD survey of civilian nonveterans, 16–21 years of age. Adjusted for unemployment.
† Incomes interpolated from data for ages 20–21 and 25–29.
‡ Median total incomes taken from *Statistical Abstract of the United States, 1966*, Table 157, p. 115. Figures were adjusted for unemployment rates of 2.8 percent for males 25 and older and with more than 8 years of education, and 4.7 percent unemployment for males with less than 8 years of education.

budget must be increased by $4 billion per year to obtain a voluntary force of 2.65 million men. The budgetary cost of moving to an all-volunteer force would be even higher if one considered the transitional period and additional retirement benefits.

Turn next to the financial cost of the armed forces to the economy as a whole. In both mixed and voluntary forces, 2.65 million men are allocated to maintaining the defense of the nation and are thereby kept out of the civilian labor force. The alternative cost of the armed forces is the value of civilian goods and services that could otherwise have been produced by them. An ideal measure of this cost would require estimates of the marginal value products of men in military service. An approximation can, however, be obtained by assuming that civilian incomes are equal to marginal value products. Median civilian incomes in 1964 classified by age and educational attainment and adjusted for unemployment are shown in Table 8. If servicemen were relocated to the civilian sector, it is assumed that they could earn the same incomes as civilians of similar ages and educational attainment. Two pieces of evidence suggest that these incomes are too low. First, the median incomes of veterans were about 2.5 percent higher than incomes of all civilians.[21] Second, a DOD survey revealed that prior civilian earnings were somewhat higher for men in the upper mental

[21] A comparison of median incomes in 1959 revealed that veterans typically earned more than all males in the civilian labor force. Veterans 25–34 years of age with 12 years of education earned 2.3 percent more than all males. An opposing pattern is observed for veterans of World War II who were over 45 years of age in 1959; there, the veterans earned less than all males. These income estimates can be found in, *U.S. Census of Population 1960*, "Earnings of Total Civilian Male War Veterans in Experienced Labor Force in 1959," Table 16 PC (2), 8C. "Earnings of Males 25–64 in Experienced Civilian Labor Force in 1959," Table 1, PC (2) 7B.

groups. Since the mental distribution of servicemen is higher than that
for the entire population (especially when education is held constant),
their alternative civilian earnings should also be higher. The use of
median civilian incomes thus imparts a downward bias to the alterna-
tive cost of labor resources allocated to the armed services.

If the mixed force described in Table 2 had earned the civilian in-
comes of Table 8, they would have received an aggregate annual
income of $13,041 million. The corresponding financial cost for the
voluntary force of Table 7 was $14,233 million or 9.1 percent greater
than that of the mixed force. Although the voluntary force has more
men with less than twelve years of schooling, it also contains more
older men, and on balance, the age effect outweighs education.

These cost estimates do not take account of personnel turnover. In
the mixed force, larger fractions of an age class enter active military
service for shorter tours of duty. The critical question is, are civilian
incomes in later life reduced by short (two to four years) tours of ac-
tive duty? The relationship of income to age (given education) can be
explained in two ways. One is that older men have acquired on-the-job
training which is reflected in higher incomes. The second argues that
age is a proxy for maturity and stability which commands higher
income.[22] If the first explanation is correct and if military training is
not a perfect substitute for civilian job experience, the financial cost of
the mixed force must be increased to reflect the cost of postponing ci-
vilian job training for more members of an age class.

The financial costs of $13.0 and $14.2 billion for mixed and volun-
tary forces are only slightly larger than the payroll cost of $12.7 billion
for the mixed force. Military pay is, however, considerably below al-
ternative civilian incomes for men on their first tours, while men in the
career force receive slightly higher incomes than their counterparts in
civilian life. It should be emphasized that these financial costs disre-
gard the occupational preferences of individuals, some of whom are
involuntarily inducted into the mixed force. They simply provide a
measure of technical efficiency in terms of the value of goods and ser-
vices which the economy relinquishes to maintain a standing army.

III. The Full Economic Cost of the Draft

The full economic cost embraces the principle that equalizing in-
come differentials are properly included in the opportunity cost of
acquiring men for military service. A simple theory of occupational

[22] Becker argues that the age profile of income is mainly attributable to investment in
human capital via on-the-job training. See G. S. Becker, *Human Capital* (Columbia Univ.
Press, 1964). The second thesis could be rationalized by my theory of fixed employment
costs. Older men are more likely to be married and possess other attributes which are
correlated with low labor turnover. If hiring and initial training costs are large, it be-
hooves the firm to offer higher wages to men with longer expected periods of employ-
ment. See, W. Y. Oi, "Labor as a Quasi-fixed Factor," *J.P.E.*, Dec., 1962.

choice along the lines of Marshall[23] provides a useful analytic frame-
work. The economy can be imagined to consist of two industries:
military and civilian sectors. Since pay cannot be separated from con-
ditions of employment, occupational preferences (utilities and disutili-
ties of the job) necessarily affect supplies of labor. Suppose that an
individual, A, could earn an income C in the civilian sector while cur-
rent first-term military pay is M_o. If military and civilian pay were
the same, $M_o = C$, an individual with an aversion for service life
would elect to remain in civil life. Military pay could, however, be
advanced by enough to compensate A for his disutility, thereby at-
tracting him into the military sector. There is, in principle, some
minimum supply price M with its accompanying equalizing differen-
tial δ such that A would be indifferent between employments in the
two sectors when $M = (1 + \delta) C < M_o$, A would volunteer for mili-
tary service.

Occupational preferences can thus be expressed in terms of equal-
izing differentials δ. If $\delta > 0$, the individual must be compensated
before he would volunteer. It is possible that some men prefer military
service in which case δ would be negative. The supply of men to the
armed forces is then determined by the joint frequency distribution
of alternative civilian incomes C and equalizing differentials δ. At cur-
rent military pay, the voluntary supply of recruits consists of those
individuals for whom $M_o > (1 + \delta)C$. If pay is advanced, the armed
forces could attract men with higher civilian incomes or with greater
aversions (meaning larger values of δ) for service life. The comple-
ment supply curve given by equation (1) implies that over the rele-
vant range, the frequency distribution of occupational preferences, δ,
follows an exponentially declining function.

The draft imposes costs on men in the armed services in at least
three ways. First, more men from an age class are demanded by the
armed forces under a draft because of the high turnover of draftees
and reluctant volunteers. Second, some men are involuntarily drafted
while others are coerced to enlist by the threat of a draft without being
compensated for their aversion to military employment. At sufficiently
high levels of military pay, all of these reluctant service participants
could, in principle, have been induced to volunteer. Finally, the true
volunteers who would have enlisted irrespective of the draft law are
denied the higher military pay that would prevail in a voluntary force.
First-term military pay can be kept at low levels because the draft as-
sures adequate supplies of initial accessions.

In the years ahead, FY 1970-75, it is projected that only 55.3 thou-
sand men will be drafted each year. If the DOD survey responses of
Table 4 are valid, another 153.7 thousand annual accessions to enlisted

[23] Alfred Marshall, *Principles of Economics* (8th ed., Macmillan, 1952), pp. 547-70.

ranks can be classified as reluctant volunteers. These projected annual flows of reluctant participants are distributed by education and age at entry in Table 9. However, all 209 thousand reluctant entrants will not be demanded by a voluntary force which is likely to enjoy considerably lower personnel turnover. In fact, the projected annual flow of 263 thousand true volunteers would have to be increased by only 75 thousand to sustain a voluntary force of 2.65 million men: 55 thousand in the Army and 20 thousand in the other services.

To measure the economic cost of the draft, one must know the minimum supply price M at which each draftee and reluctant volunteer could be induced to become a true volunteer. Such data are simply unavailable, and some simplifying assumptions must be invoked to arrive at an estimate of this cost. It seems reasonable to suppose that the 153.7 thousand reluctant volunteers who entered as regular enlisted men in preference to being drafted have less aversion to military service when compared to draftees and men who never enter service.[24] Indeed, if the draft were abolished and military pay advanced, I suspect that most of these men would become regular volunteers. A lower bound to the economic cost is thus obtained by assuming that these reluctant volunteers have the lowest minimum supply prices. According to the complement supply curve for enlistments to all services, first-term pay must be raised by a factor of 1.88 to attract an additional 153.7 thousand recruits.[25]

The exposition is facilitated by referring to the supply curve of Figure 1. If the draft is eliminated with no pay changes, the annual

[24] It has been alleged that many reluctant volunteers are prompted to enlist in preference to being drafted because of the uncertainties about whether they will be drafted, or if so, when. In his doctoral dissertation, David Bradford argues that men who become reluctant volunteers because of this uncertainty are those for whom the relative cost of later military service is the greatest. He further argues that the absolute cost of active military service is immaterial. I believe that his argument is specious because he considers only a dichotomous choice. The gist of his simple model can be put as follows. Let C_0 be the subjective cost to the individual of active military service now, while C_1 is the cost of entry at a later date and older age. If p is the subjective probability that he will be drafted later, this individual would prefer to enlist now if $C_0 < pC_1$ or $(C_0/C_1) < p$. Hence, only the ratio or relative cost of later entry, (C_0/C_1), matters. There is, however, a third option available to every draft liable youth; namely, securing a deferment. This may entail a cost C_d especially if he must become a father or minister, or if he must pursue an occupation which bestows an occupational deferment. There are many reasons to suppose that the cost of acquiring a deferment, C_d, is independent of C_0 and C_1, the subjective cost of actually entering military service. Hence, if $C_d < C_0$, the individual would secure the deferment and never enter military service. Moreover, if C_d is the same for all individuals, Bradford's analysis applies only to men for whom the cost of active military service C_0 is greater than C_d. For the details of Bradford's analysis, the reader is referred to, David F. Bradford, "The Effects of Uncertainty in Selective Service," Technical Report No. 144 (Inst. for Math. Studies in the Soc. Sci., Stanford Univ., 1966).

[25] The complement supply curve, equation (1), for enlistments to total DOD revealed an estimate of β of .315 with a standard error of .087. In the absence of a draft, the enlistment rate (Mental Groups I to III) was estimated to be 27.8 percent of the qualified civilian labor force in an age class. In deriving the supply curve shown in Figure 1, I set β equal to .402, the point estimate plus one standard error. The elasticity of this supply curve at the initial enlistment rate was +1.04. Because of the upward bias in the supply elasticity, my procedure understates the magnitude of the economic cost.

TABLE 9

PROJECTED ACCESSIONS TO ENLISTED RANKS UNDER A CONTINUED DRAFT, FY 1970–75
(By Age at Entry, Education, and Draft Motivation)

| | Total | Years of School Completed | | | | | |
		0–8	9–11	12	13–15	16 and Over
Voluntary enlistments under a continued draft						
17–19 years.................	328.5	15.0	107.2	188.0	18.3	0
20 years and older...........	88.2	1.4	12.9	42.8	25.8	5.3
All ages...................	416.7	16.4	120.1	230.8	44.1	5.3
Reluctant volunteers						
17–19 years.................	102.4	3.0	23.0	68.3	8.1	0
20 years and older...........	51.3	0.4	5.3	24.7	17.8	3.1
All ages...................	153.7	3.4	28.3	93.0	25.9	3.1
Inductions..................	55.3	4.3	15.2	21.9	10.5	3.4
Reluctant military service participants						
17–19 years.................	102.4	3.0	23.0	68.3	8.1	0
20 years and older...........	106.6	4.7	20.5	46.6	28.3	6.5
All ages...................	209.0	7.7	43.5	114.9	36.4	6.5

SOURCE: Derived from Table 4.

FIGURE 1.
SUPPLY CURVE OF VOLUNTARY ENLISTMENTS TO TOTAL DOD

supply of voluntary enlistments is projected to be around 263 thousand men at a first-term pay $M_0 = \$2,500$. If pay is increased by a factor of 1.88 to $M_1 = \$4,700$, the annual supply of recruits climbs to 416.7 thousand men; i.e., from OA to OB accessions. The reluctant volunteers (the line segment AB) enlist at the lower pay M_0 in order to avoid being drafted. The difference between their minimum supply prices and the current first-term pay M_0 represents an implicit tax which is borne by these men. The aggregate annual cost for the 153.7 thousand reluctant volunteers is thus given by the area of the triangle, $DB'E$, or $\$141$ million. This estimate tacitly assumes that each reluctant volunteer is compensated in a discriminatory fashion without compensating the true volunteers. If, however, pay were raised to $\$4,700$ for all recruits including true volunteers, the annual cost of the draft is increased by the additional amount M_0DEM_1 or $\$917$ million.[26] The lower annual cost of $\$141$ million which excludes rents represents an implicit tax levied against reluctant volunteers who

[26] In competitive labor markets, these rents are always included in the additional cost of acquiring larger supplies of labor. If college professors could be conscripted and compelled to work, or if the economic rents of teaching versus other occupations could be eliminated, the budgetary cost of higher education could be greatly reduced.

were coerced to enlist by the draft liability. In a sense, each reluctant volunteer pays, on average, an implicit tax of $915 in each of the 3.5 years of his first term of service. If the point estimate of β had been used in these calculations, the aggregate annual cost, $DB'E$, is estimated to be $192 million.[27] Since each regular enlistee serves for 3.5 years, the total tax (excluding rents) borne by the reluctant volunteers in an age class is conservatively estimated to be $493 million; the best estimate is $672 million.

.The economic cost of conscripting men into military service is harder to assess. The Selective Service System does not attempt to draft men with the least aversion for military life. The supposition that draftees were next in line above the point E in Figure 1 is less plausible than in the case of reluctant volunteers. However, a lower bound estimate is again obtained by assuming that draftees had the smallest equalizing income differentials and hence the lowest minimum supply prices. In Figure 1, first-term pay must be raised from $M_1 = \$4,700$ to $M_2 = \$5,900$ to attract the 55.3 thousand draftees (the line segment BC) on a voluntary basis. If each draftee is compensated in a discriminatory fashion, the implicit annual tax which is borne by involuntary draftees is given by the area $EB'C'F$ or $175 million.[28] Since the average active duty tour for a draftee is about 1.9 years, the total implicit tax for draftees in an age class is $333 million.

Each reluctant volunteer and draftee could, in principle, have been induced to enter active military service on a voluntary basis. The draft, however, compels some to serve while others are coerced to enlist at military pay scales which are below their minimum supply prices. The difference between minimum supply price and current first-term pay is simply an implicit tax—the economic cost of active military service for reluctant service participants. A lower bound estimate of this cost (for those who serve in enlisted ranks) is derived from the area $DC'F$ and is approximately equal to $826 million for reluctant participants in an age class. If the least squares estimate of the supply of voluntary enlistments were used in the supply curve of Figure 1, I obtain the middle estimates in the last column of the following summary table:

[27] If the value of β is reduced from .402 to its least squares point estimate, .315, the elasticity of supply falls from 1.04 to 0.82. Using the point estimate, the necessary first-term pay M_i to attract OB volunteers rises from $4,700 to $5,600.

[28] Since the draftee serves for only two years, his annual military income of $2,100 is less than the first-term pay of regular enlistees, $M_0 = \$2,500$. Hence, the annual cost is slightly greater than the area $EB'C'F$ in Figure 1. If β is set equal to .315, first-term pay must rise to $M_2 = \$7,450$ to attract the draftees on a voluntary basis. In this event, the annual cost climbs to $243 million, and the aggregate cost for the draftees in an age class rises from $333 to $462 million.

ESTIMATES OF THE IMPLICIT TAX ON RELUCTANT SERVICE PARTICIPANTS

	Low Estimate $(\beta = .402)$	Middle Estimate $(\beta = .315)$
Annual first-term pay		
M_0	$2,500	$2,500
M_1	4,700	5,600
M_2	5,900	7,450
Annual cost excluding rents (millions)		
Reluctant volunteers $(DB'E)$	141	192
Draftees $(EB'C'F)$	175	243
Aggregate cost for an age class, excluding rents (millions)		
Reluctant volunteers	493	672
Draftees	333	462
Total...............................	826	1,134

The economic cost or implicit tax placed on men who were coerced to serve by the draft provides a lower bound estimate of the opportunity cost of acquiring enlisted men. The estimates shown in the preceding table are biased downward because the men who bear the cost are assumed to be those with the lowest supply prices in the absence of a draft. These estimates also neglect the rents that would have been paid to true volunteers in a competitive labor market. Under a draft, we not only tax the reluctant service participants but we also prevent true volunteers from collecting these rents. The full economic cost which includes these rents is estimated to be $5,364 million when I use the supply curve of Figure 1.

It is of some interest to separate the economic cost of the draft into (1) the loss of alternative civilian income during active duty service and (2) the net sum of equalizing income differentials to overcome aversions for service life. If the median 1964 civilian incomes of Table 8 are applied to the probable distribution of reluctant participants, Table 9, the average alternative civilian incomes are:

ESTIMATED ANNUAL CIVILIAN INCOMES OF DRAFTEES AND RELUCTANT VOLUNTEERS

	Annual Flows (Thousands)	Civilian Income	Military Income	Ratio
Reluctant volunteers ..	153.7	$3,450	$2,500	.72
Draftees	55.3	3,810	2,100	.55
Total................	209.0	3,545	2,400	.68

The differential between alternative civilian and military incomes obviously varies, being larger for older, more educated men. The infrequent college graduate who is involuntarily inducted can expect a financial loss of over $3,000 a year. The aggregate financial cost to members of

an age class (the difference between alternative civilian and current first-term pay) is estimated to be $691 million. This financial cost results from the abnormally low levels of first-term pay and is independent of the occupational preferences of reluctant service participants.

In this paper, the economic cost has only been estimated for men who serve in enlisted ranks because of the draft. A fuller analysis would include the costs borne by draft-motivated accessions to officer ranks and to reserve and National Guard units. The available evidence suggests that these men (especially those who enlist in Guard and reserve programs) have high aversions for military service.[29]

In addition to the direct costs borne by those who ultimately serve in the armed forces, the draft allegedly creates other indirect costs which derive from the mechanics of the selection process. Under the current Selective Service System, a youth can remain in a draft-liable status for seven and a half years. There is some evidence which suggests that employers discriminate against youths who are still eligible to be drafted.[30] The youth who elects to wait and see if he can avoid military service is likely to suffer more unemployment. He may be obliged to accept casual employment which does not provide useful job training for later life. Moreover, long periods of draft liability encourage youths to pursue activities which might bestow a deferment. When married nonfathers were placed in a lower order of call in September, 1963, it was followed by small increases in marriage rates of males in the draft-liable ages. It is also alleged that the draft prompts men to prolong their education or to enter occupations which grant deferments. These costs which derive from the uncertainty of the present draft are, in my opinion, small when compared to the direct economic costs incurred by those who are involuntarily inducted or who become reluctant volunteers.

If the current draft law is extended into the decade ahead, it is pro-

[29] According to the 1964 DOD survey, draft-motivated enlistments accounted for 70.7 percent of total accessions to reserve and Guard components. Moreover, the percentage of reluctant volunteers was higher for individuals with higher alternative civilian earnings as revealed by the following excerpt from the *House Hearings* (p. 9956).

Annual Civilian Income	Percentage of Draft-Motivated Enlistments
Less than $2,999	54.1
$3,000-$4,999	71.6
$5,000-$7,499	72.1
$7,500 or more	82.2

[30] A survey of college placement offices in 1964 indicated that 48 percent of employers placed restrictions on draft-liable males. According to the DOD survey of 22-25 year old civilians, 26 percent with draft classification I-A said that they had difficulty in securing employment. The credibility of this type of question is challenged by the finding that 17 percent of men with dependency deferments, III-A, also stated that they had difficulty securing employment because of their draft liability. The percentages having "difficulty" climbed with age and educational attainment. Further details of these surveys can be found in *House Hearings*, pp. 10008-10010.

jected that only 38.5 percent of qualified males will be required to staff a mixed force of 2.65 million men. Since the draft assures adequate supplies of initial accessions, military pay can be kept at artificially low levels. Many servicemen on their first tour can correctly be called reluctant participants who pay substantial implicit taxes because they were coerced to serve. A conservative estimate of the economic cost (excluding rents) is $826 million—the amount of compensation which would have been demanded by these men to enter on a voluntary basis. If all recruits received the first-term pay needed to attract the last draftee, the opportunity cost of acquiring new accessions would exceed $5.3 billion.

An all-volunteer force offers a polar alternative to the draft. With its lower personnel turnover, a voluntary force of the same size could be sustained by recruiting only 27.5 percent of qualified males. The budgetary payroll cost would, however, have to be raised by $4 billion per year.

In closing, it should be emphasized that the figures appearing in this paper represent my estimates. The two crucial ingredients are the supply curve of voluntary enlistments in the absence of a draft and projected gross flow demands which are determined by personnel turnover. Complement supply curves were estimated from cross-sectional data on voluntary enlistment rates.[31] The retention profiles which were used to derive gross flow demands for an all-volunteer force generated an age structure of the force which closely approximates that of smaller professional armies in Canada and the United Kingdom. In the light of the data examined, I am reasonably confident of my cost estimates, at least for the assumed force strength of 2.65 million men.

If peacetime military requirements necessitate larger active duty forces, all costs necessarily climb. To sustain a force of 3.3 million men on a voluntary basis, the gross flow demand for new accessions rises by roughly 30 percent. Estimation of the pay increase to achieve this larger force involves an extrapolation of the supply curve beyond the range which I consider to be meaningful.[32] The high budgetary cost

[31] Such data on a regional basis were only available for 1963. The recent enlistment experience is strongly influenced by the Vietnam build-up and the high draft calls. In any move toward a voluntary force, I would strongly recommend an on-going research effort on the determinants of the supply of volunteers as well as on manpower utilization practices of the armed forces.

[32] The confidence interval for predictions from a regression equation becomes extremely wide as one moves outside the range of the sample observations. Hence, the following estimates should be regarded as only indicative of orders of magnitude. In order to sustain a voluntary force of 3.3 million men, Army enlistments must be increased from 90.3 to 185 thousand per year; the latter gross flow demand of 185 thousand assumes a steady state. According to the complement supply curve, first-term pay would have to be raised from $2,500 to $6,350 to attract sufficient recruits. If part of the pay could be given as initial enlistment bonuses, the defense budget for active duty military personnel would be increased by $8 to $10 billion per year.

DISCUSSION

GEORGE H. HILDEBRAND: There can hardly be a better subject for economic analysis than comparative methods for the recruitment of military manpower. The problem has both micro- and macroeconomic aspects. It involves both allocative and distributive effects. It raises questions of wage structure and of labor mobility. It combines the tasks of formal analysis and of statistical estimation. Above all, given the highly controversial ambient of issues surrounding military recruitment in the United States today, the economist who ventures into the field has both the challenge and the opportunity to separate clearly his scientific findings from the normative aspects of the question.

On all these counts, then, it is indeed surprising that until recently so little professional attention has been paid to the economics of the recruitment and use of military manpower. The two papers by Oi and by Altman and Fechter represent a real start toward opening up the discussion.[1] Since time is short and the two papers overlap considerably, I shall concentrate upon the one by Oi.

Oi takes as a datum 2.65 million men for the chosen level of the force in being. All of his calculations rest on this assumption. Undoubtedly this was a procedural necessity, but it must be observed that the present level is 3.3 million men, which must mean that the cost of going over to a fully voluntary system will be considerably higher than his figure of $4 billion in increased payroll costs.

Oi's first step is to determine the drop in the current accessions rate upon adoption of an all-volunteer method. This rate includes those men who are drafted and who are what he calls "reluctant volunteers." Oi expects that on a volunteer basis the rate would fall, because the necessary adjustments in pay levels should increase reenlistments and lower turnover.

Oi's second step is to estimate the rise in military pay levels required at the starting step and beyond to produce the desired adjusted accessions rate. In essence, the problem here is to determine the probable elasticity of manpower supply, given various ratios of military pay to alternative median civil incomes. Using cross-section data, Oi obtains an initial elasticity of supply of 1.36, which is rather high. From this supply equation, Oi concludes that if first-term pay were raised by 68 percent, that is, from $2,500 to $4,200, with a further increase of 17 percent for the career groups above, the desired accessions rate could be achieved. Once the transition to a steady state were completed, Oi thinks that the all-volunteer system would raise payroll costs as he calculates them from $12.05 billion currently to $16.103 billion finally, or by about $4 billion, to maintain a force in being of 2.6 million men. For the current higher level of 3.3 million men, the added cost would be much greater, and one wonders whether the estimated supply elasticity would not also be

---

[1] Because I did not receive the Fisher-Morton paper until three days before this meeting, I am unable to include it in my remarks.

somewhat lower. Assuming no disadvantageous qualitative effects on the input of recruits, the price might still be well worth paying, for it would permit us to escape all of the rigamarole and inequities of the present selection and deferment system.

But there are questions still to be asked, for the reliability of Oi's figure depends upon the answers to them. One concerns the estimated drop in the accessions rate upon shifting to an all-volunteer system. Oi states that abolition of the draft would cut the present rate by 21 percent, which would drop by an additional 17 percent through loss of the so-called "reluctant" volunteers. He also expects the needed accession rate itself to decline somewhat, through lower turnover. What we need to know is the reliability of the estimates for all three factors. What is the range of their standard errors?

The second main question concerns the stability of the estimated supply curve. Rightly in my judgment, Oi makes the supply of military manpower, as with that for any other occupation or profession, to depend upon relative rather than absolute earnings—in this case, military pay in relation to average civilian income or opportunity cost. Clearly, then, if through time civilian earnings rise while military pay is held at the new levels, the supply function supposedly would shift inward—meaning that military payroll costs must be raised still further to preserve the desired accessions rate. Further, as Fechter and Altman have established from their analysis of time series, the supply function will be affected by the applicable civilian rates of unemployment for the male manpower groups involved. If these rates tighten still further, there would be adverse supply effects for the military. Finally, I suspect that both the slope and the position of the supply function will shift with changes in public opinion toward international tensions and conflict. Men are more likely to volunteer at a given rate of pay when they are convinced that the nation's security is threatened than when they are doubtful that such is the case—given any prescribed level of forces in being. In short, nonpecuniary influences will affect the supply function.

Let us turn briefly to a larger theme posed by both the Oi and Fechter-Altman papers. Although I confess to finding the voluntary approach preferable to a coercive one—be the latter a draft with exemptions or a lottery—I do think that the scale of forces required may well be decisive for the choice among alternative methods of recruitment. It seems to me that the voluntary system would work, and work better, in two situations: (1) for providing a standing army of stable size during times of genuine peace or even of continuing cold war; and (2) for providing a somewhat larger force, say, to cover normal security functions and to fight limited wars on the frontiers, such as we are now undertaking in Vietnam. However, I am not convinced that the voluntary method can work where the level of required forces approaches that relatively comparable with the peak of World War II, in which 11 million men were in service. Given such a scale of economic mobilization and a probable accompanying panoply of direct controls, is it feasible to think of relative real income as a method for making voluntary procurement work effectively in such circumstances? At the very least, this issue is well worth more serious exploration.

A related question concerns the administration of the pay structure itself under a voluntary system. Suppose that the estimated levels fail in fact to produce the desired rate of accessions. Or that a change of requirements raises that rate substantially relative to the flow actually forthcoming. Is the Department of Defense to be given the fiscal flexibility needed to experiment with supply conditions? Should it be free to raise pay rates, or to introduce enlistment bonuses? Should it retain reserve authority to reinstate the draft? Or would it be enough for the Department to act like a private employer in the same circumstances; that is, to adjust its quality standards for admission upward or downward as conditions vary between excess demand and excess supply? And what about internal revisions of the pay structure, to manipulate the reenlistment rate or to divert career personnel to more urgent uses? Clearly, the Department must have considerable autonomy in "wage" administration if it is to strive for a set of equilibirum values affecting manpower; that is, if it is to procure the desired number of new entrants of requisite quality, and if it is to achieve a distribution of its manpower in the right amounts for alternative uses.

Let me turn now to a brief discussion of the broader economics of the problem, partly because I think that the conceptual framework still needs development, despite the valuable contributions of both papers at hand.[2]

First, the present draft system is a coercive method for the procurement of real resources by the state; in this case, the services of human beings. On narrow grounds, its rationale would be that it is the only reliable way to obtain the needed manpower. This may well be true in some circumstances, but the case surely needs careful examination. More particularly, even if the case turns out to be sound, it implies no support for one of the decisive characteristics of the draft itself: that for many men the military rate of pay is below their opportunity cost earnings in civilian life. Why should servicemen be required to pay this special tax for the benefit of the rest of the taxpayers? In principle, selective service could be coupled to payment of equivalent market wages. Not to do so is to hold down the cost of the military payroll at the expense of those men in service whose transfer prices or equalizing differentials exceed their military earnings. On equity grounds this tax ought to be borne by the civilian community.

Second, because military "labor" is now to some extent undervalued, the armed services are given a false signal by the price system: they are encouraged to use labor more intensively relative to capital than is justified by the real state of relative factor endowments in the economy as a whole. In consequence, it pays to hoard labor, to use it wastefully, and to adopt capital-to-labor ratios that are too low. Turnover rates are also made too high, and these add to recruitment costs while also lowering overall efficiency.

Third, below-market military rates of pay plus the draft probably yield an overall force of lower quality than could be had under a voluntary system. The reason is twofold: not enough of the "right" men will volunteer, while the haphazard network of exemptions tends to concentrate selection among less

[2] Here I shall depend in part upon a very stimulating working paper by W. Lee Hansen and Burton A. Weisbrod, "Economics of the Military Draft" (unpublished, Oct. 28, 1966).

fitted recruits—less fitted in general educational preparation and career poten-
ial, particularly for promotional advancement. Private employers who seek
high efficiency, low turnover, and promotable potential have long ago dis-
covered the advantages of competitive and even supercompetitive starting
rates, of rationally designed internal wage and salary structures, and of fringe
benefits as retention devices. Even if a persuasive case can be made for perma-
nent reliance upon the draft, it is time to consider the manpower economics of
the military pay system. Even more, we ought to strive hard to develop a fea-
sible voluntary system to replace the draft. To refuse to do so on the ground
that the result would be a mercenary army makes no more sense than to say
that there is something wrong in using a mercenary labor force either to make
steel or to police the streets. The method of free exchange has enormous in-
trinsic superiority over direct and centralized control and a command system
of economics—just for getting things done well. Why should a force in being
of 3.3 million men be an exception?

PAUL A. WEINSTEIN: These research papers mark a significant genesis in
the economic analysis of military manpower which is long overdue. It is
hoped that we have unlatched the gate that separates the academic from the
military establishment. Here is one area where we are deeply indebted to
the military mercenaries of this panel. The techniques now under development
will have considerable bearing on military decisions at some future time.

It is unfortunate that on the basis of the material at hand, we are willing to
warrant our policy prescriptions with such certainty when even the economic
consequences lie in a murky penumbra. The results and basic data in two pa-
pers are at variance. While Oi refers us to Altman and Fechter on estimating
the supply functions, we observe marked differences. For example, the Oi esti-
mate of the price elasticity of supply based upon Altman and Fechter is 1.36,
while Fechter's is 1.17. The base pay with allowances in Oi's study is $2,500,
while $3,415 in the other. Thus, if one applies the 1.8 equivalent factor to the
Altman-Fechter base pay of first-term enlistees, the required pay jumps to the
pay of urban police and firemen—a comparison which is not inappropriate.
Discrepancies such as these and assumptions about the force's quality give
testimony to the cost range of a volunteer army—from $4 billion to $17 bil-
lion. Better comparisons may be forthcoming from international experience.

I think one of the most important salients is found in the research of Pro-
fessor Walter Oi. Here is a positive attempt to estimate the real allocative
effects of the military on the overall economy and the allocative, as well as
budget, costs of alternative types of manpower procurement. The questions of
the size and the burden of the draft's implicit tax are central to evaluating the
resource implications of the current draft. It is unfortunate, however, that in
the presentation of these costs no significant attempt was made to consider
benefits.

The difference between the military and the alternative pay rates is consid-
ered as a tax payment with a transfer benefiting the non-taxpayers. In these
estimates, Oi is much more thorough than Altman and Fechter. An alternative
interpretation of the differential between military pay and alternative pay re-

quires a significantly different time frame. This interpretation would look upon the differential as an investment, albeit an involuntary one. This investment is partly in income generating capital; that is, military experience is productive in the market sector. An unknown portion of those passing through the service receive their current benefits and a reward in the form of capital which has economic value. My research indicates that conscripts with no economic interest in the military tend to capitalize on the experience. The degree to which this investment serves as an offset to manpower misallocation is an empirical problem deserving study. For example, the urban Negro or rural white who becomes a policeman, console operator, or mechanic as a result of the military is advantaged by the military, and this addition should be evaluated in estimating the draft's cost. As one who is currently conducting research in this area, I am aware of the theoretical and statistical difficulties in measuring these magnitudes, but suggest that Professor Oi's dismissal of this is serious and the importance is noted in the yet unpublished paper of Professors Weisbrod and Hansen.

The other type of investment forced on the draftee is more like the purchase of a consumer's durable. It is a mix of products that we desire for our society. In effect, the draft is converted to eminent domain of human capital. It is analogous to the use of eminent domain in the compulsion and pay scale used in our jury system. Should we adopt a volunteer jury system with pay as the exclusive variable to secure jurors? Many would reject this because it would result in a transform of the product, a fair trial by peers. Similarly, one may question whether the use of pay as the sole variable in the military supply may transform the military.

On this point, both papers raise problems requiring considerable quantitative research. Oi notes that under a volunteer army the average age of the military is increased and the average education level of the force strength is reduced. Altman and Fechter give sharper focus to this showing that the higher mental aptitude groups (I-III) and the more educated are less willing to enlist with the advantaged having a low price elasticity to the military. Their estimates, with the peculiar $5\frac{1}{2}$ percent general unemployment rate, project a diminution in the military manpower inventory. The question they pose is how many bodies, abstracting from any qualitative variable, will be forthcoming to replace the bodies exiting the military. Accepting the analysis of regional tastes for the military, which may be a proxy for relevant unemployment rates or education levels and its surrogate in the mental test, we are left with the conclusion that the volunteer military will be composed of low aptitude, ill-educated, and older soldiers. The problem is accentuated as we reduce the unemployment rate or tie enlistment and reenlistment rates to age and area specific unemployment rates.

If the technology in the military continues to alter and become more complex, we must assume under the OFA analysis that the productivity of the military both in current performance and in expansibility potential is invariant to the age, education, and aptitude of the labor component. The analysis clearly implies the elasticity of substitution is infinite between all individuals in the military regardless of quality, but not time in service. If this is correct,

our current military manpower judgments are in error. This notion is inimical to human capital theory. They do not accept the military production function but alter it radically and improperly. If their analysis is not correct, then the real output of the military depends upon a qualitative mix in the manpower input. While we may be able to get the same "bang" out of inefficient people as their ascribed betters, we would undoubtedly need them in greater numbers, hence at increased cost. It is obvious more must be done on the manpower mix required per unit bang, as well as work on the absolute size of the military establishment.

Parenthetically, one might note the defense of the volunteer army based on a lower cost of turnover. The low turnover rate implies that our less able, ill-educated, older military is as adaptable to new technology, hardware systems, and techniques as a younger group of individuals. I suggest that we should not try to minimize throughput cost, but try to optimize some level of throughput which allows us to absorb new hardware systems and maintain the capacity which is required to meet potential manpower needs at some reasonable cost.

It is interesting that throughout the statistical work, wherever there is a bias, the bias is in favor of a low quality input.

Altman and Fechter in their analysis of the officer supply have estimated the income to college graduates five years out of service. This five-year span would exclude high-income earning professionals and graduate students. They edit their figures—and there is justification for it—to eliminate those schools with the median income in an open-ended category. One must assume this open-ended category is a high-income group. Hence, they probably eliminated in editing the data schools such as Harvard, Yale, Princeton, and the like. In fact, we are left with schools of lesser stature. If we assume the graduates of these other institutions are not quite as able as graduates from schools in the open-ended category, we are left with an indication of a low cost. However, the low cost is only an expression of a possibly low quality that would be coming into the military.

The contribution of Fisher and Morton is sizable for its technique in evaluating the military production problem inclusive of human capital. It is also germane to the cost of eliminating compulsory service by recognizing the quality mix problem. While alternative techniques to estimate the marginal rate of substitution may improve the results or by adopting some cardinality through externally provided discounted "kill" rates might yield improvement, the immediate impact of their work is to suggest a differential pay policy to alter turnover rates (what all this may do to shifting the production function by affecting morale is an unknown). Consideration of education and aptitude jointly with "time in service" as a proxy for productivity, will alter the quantitative measures of FTRR. The volunteer system will affect costs by altering the marginal rates of substitution through the erosion of skill at all ability levels. These costs should be evaluated in estimating the costs of altering the procurement system.

I am not suggesting that we may currently overbuy quality in the military. I am not suggesting that the volunteer army may not be an appropriate solution to the manpower problem. I am not suggesting the implicit tax may still

outweigh significantly the explicit budgetary cost that we have in a draft, or that we may raise the military wage of first-termers to reduce the tax. What I am suggesting is the papers prepared under the auspices of the Defense Department by these heroic scholars have produced what is at best a pig in a poke. Both for economic and dietary reasons, this merchandise is unacceptable.

HAROLD WOOL: By any conventional economic criteria, one of the largest and most consistent areas of labor shortage in this country has been a shortage of manpower for military service. The very fact that a draft law has been necessary almost continuously since the end of World War II is evidence that conventional labor market mechanisms under the existing levels of pay and incentives have been insufficient to attract a sufficient number of volunteers to military service.

The initiative for studying the economics of military manpower supply, in the face of this record, has come quite properly from the Department of Defense. In 1964, the Department of Defense initiated a major study of the draft, designed—among other things—to assess the feasibility of meeting military manpower needs in the coming decade on a completely voluntary basis. Those concerned with formulating our long-range manpower policies turned to the economists for answers to the following questions: Will it be economically feasible to recruit and maintain peacetime military forces of the size required in recent years on a completely voluntary basis? And, if so, how much would it cost?

These questions appeared simple. The policy-maker would have welcomed simple answers to these questions. It would have been tempting to oblige by submitting simple unqualified responses to these questions. In view of the present state of our knowledge, I believe, however, that this would have been a serious error.

First, it is important to recognize that the subject matter we are dealing with is not a theoretical exercise. It is vitally related to our national security and to the ability of our armed forces to meet their commitments in future years. It also directly affects the welfare of millions of young men in our country.

In this context, let us consider the nature of the information available to arrive at these estimates. In large part, the estimates of the responsiveness of military recruitment to pay, which served as the basis for both the Altman-Fechter projections and those of Walter Oi, rested on a cross-sectional analysis of geographical differences in enlistment rates by census geographical region in relation to differences in civilian earnings and unemployment rates. The statistical foundation for this analysis, therefore, consisted only of nine observations, i.e., adjusted "voluntary" enlistment rates in 1963 by census region, which were correlated with regional earnings and unemployment rates. The implicit assumption in both papers, moreover, is a nearly instantaneous adjustment of military manpower supply to adjustments in first-term pay.

These estimates and assumptions are inevitably fraught with great uncertainties. There is much which we still have to learn about the mechanism of

occupational choice—about the influences which condition young men to choose a military career as against other alternatives in civilian life. We do, however, know from a variety of sources that attitudes towards military service as well as to other occupational careers are often formed early in adolescence and are influenced by many cultural and sociological factors other than pay. In fact, some of our surveys have shown widespread ignorance of the true level of military compensation on the part of civilian youth. It is most unlikely, therefore, that short-range supply elasticities would approach those suggested by the cross-sectional analysis.

Our uncertainties concerning the responsiveness of recruitment to pay incentives are compounded by uncertainties concerning the effects of changes in the civilian labor market upon recruitment. Much of the empirical data available for the Department of Defense studies was developed under conditions when the national unemployment rate averaged about 5.5 percent and when the unemployment rate for teen-age youth was two to three times as high as this average. The cost estimates developed in the Department of Defense study are based on a range between a 4 percent and a 5.5 percent unemployment level. These were presented in the Fechter-Altman paper; Oi chose to base his estimates on a 5.5 percent unemployment rate assumption. Even the lower of these two average rates coincided with unemployment rates for teenage youth of about 10 percent during the past two years. A major national manpower policy objective is to reduce these youth unemployment rates still further through improved educational and training programs, improved labor market mechanisms, and other measures. We unfortunately do not have experience to assess the effects of still lower youth unemployment rates on availability of youth for military service.

For these and many related reasons, I believe that any single unqualified estimate of the "cost of an all-volunteer force," tempting though it may be, exceeds our capability at the present time. At best, these cost estimates must be stated in a broad range to suggest the limits of our knowledge and the inherent error of estimate in such projections.

COST-BENEFIT ANALYSIS FOR
GOVERNMENT DECISIONS

GOVERNMENT-INDUSTRY DEVELOPMENT OF A
COMMERCIAL SUPERSONIC TRANSPORT*

By STEPHEN ENKE

The U.S. government is expected to spend $2 billion or more to develop a commercial supersonic transport (SST) that will be safe, profitable, and available for airline use around 1974.

Among the major policy questions now being raised by the U.S. SST program are: (1) Is a U.S. SST economically justifiable? (2) Why is federal assistance necessary? (3) How much of its expenditures on the SST should the federal government recover, and through what means? (4) How can government finance an SST monopoly, and yet protect the public interest without concerning itself with aircraft prices, flight frequencies, and passenger fares? (5) Are the benefits of the SST sufficiently general to justify a federal subsidy of its development and possible manufacture?

The answers to these questions are of very broad interest because the U.S. SST may be the first of several federal govenment programs to develop products of advanced technology for commercial use.

SST's Economic Justification

The President has declared that the U.S. SST must be safe and "profitable," but what is the interpretation of profitable and how can it be estimated far in advance?

Airline managements presumably will not buy an SST unless it can earn as much on its investment cost as can be earned on the advanced subsonics of the 1970's (e.g., the Boeing 747). Practically, this means that SST prices, flight costs, and receipts must permit a return of about 20 percent before taxes and interest. (Publicized "orders" for SST's are revocable until performance and prices are known.)

What airframe and engine manufacturers must ordinarily expect to earn from a new aircraft is more uncertain. Realizations have often been far worse than expectations, and several new commercial aircraft of the past ten years have lost money for their manufacturers (e.g., Convair 880). Expectations of profit must presumably be higher the

greater are the sums to be risked and the greater the dispersion of possible financial outcomes.

If the U.S. government is initially to finance 80-90 percent of the U.S. SST's development costs, and if up to $4 or $5 billions of American resources must somehow be invested in development and manufacture before the program generates a net cash inflow from sales to airlines, an obvious economic test of the U.S. SST's justification is its ability to earn the 10-15 percent rate of return earned on an average by domestic resources employed in U.S. industry.

Whether the resource and money costs sunk in the U.S. SST's development and manufacture can earn a 10-15 percent rate of return depends upon: (1) costs of development, for airframe and engine; (2) costs of manufacture, for airframe and engine, as a function of production rate, cumulative output, and calendar date; (3) operating costs per plane mile, as a function of various performance parameters (e.g., specific fuel consumption) and permissible altitude of transonic acceleration; (4) operating receipts per plane mile, as a function of seating capacity, load factor (percentage capacity sold), and fare levels; (5) availability of supersonic passengers, as a function of number of long-haul passengers, subsonic passenger fares, willingness to pay more for less time in the air, and extent to which sonic boom nuisance restricts available routes and schedules. All of these factors remain shrouded in uncertainty.

Development costs of aircraft have often been underestimated by a factor of two—especially where the necessary state of art has yet to be attained.

SST operating profits are extremely sensitive to specific fuel consumption, lift-drag ratio (a measure of aerodynamic efficiency), and aircraft weight empty. Of its maximum gross take-off weight of maybe 650,000 pounds 10 percent or less will be payload and about 45 percent will be fuel. An X percent improvement in fuel economy, if translatable into greater seating capacity, could increase net flight receipts by perhaps $3X$ percent, proportionately increasing the selling price the aircraft can command.

Supersonics will have to compete with subsonics. Advanced subsonics may have costs per seat mile 20-30 percent below those of the U.S. SST. Airlines, unless they acquire SST's at subsidized prices, will hence buy relatively few SST's so that load factors are exceptionally high (e.g., 70 percent of seats are filled instead of the 55 percent typical of today's jets) and have to levy a surcharge (e.g., 10-20 percent) on SST tickets. No one knows how many long-haul passengers would be diverted to subsonics by such fare differentials. Estimates are that one-half may be lost.

The growth rate of passenger demand is also uncertain. Long-haul

revenue passenger miles have increased at an average of 14 percent during the past ten years, but this rate is expected by most to decline. Continued growth will presumably vary with quality of service, per capita incomes, and population growth. The lowest estimate of passenger volume growth between 1966 and 1990 is five times.

Public acceptance or nonacceptance of frequent sonic booms by commercial aircraft remains doubtful. If commercial overland supersonic flight is not permitted by the U.S. and most foreign governments, no more than 40 percent of long-haul passengers remains to be shared with subsonics on the available routes over water. Conversely, as the SST's engines are "sized" for high altitude transonic acceleration (above 35,000 feet) and fuel consumption per mile at subsonic speeds is much higher, the economics of the SST are improved if it is permitted to cross the sound barrier at lower altitudes nearer to airport terminals. Unfortunately, as SST designs increase in gross take-off weight to permit more seats and hence lower seat mile costs, the severity of expected sonic boom increases and the possibility of operating restrictions (e.g., curfews on transcontinental flights) is increased. The prudent assumption, until more conclusive predictions can be made of public reaction to boom, is that commercial supersonic flights over populated land areas will be prohibited.

Anyone who has been intimately involved in econometric assessments of SST "profitability" realizes full well that its prospects will remain most uncertain at least until prototype flight tests and possibly until regular commercial operations begin. The total market for SST's to 1990 is probably somewhere between 150 to 600 aircraft. The profitability of the program is sensitive to numbers of SST's sold. Under certain sets of favorable assumptions the overall rate of return could be 10-15 percent. But under other sets of equally plausible assumptions the return is negative. Perhaps a best guess is an overall program rate of return of 0 to 5 percent.

The simple truth, however, is that such a complex and technically advanced aircraft must be an investment gamble in its first generation.

Necessary Federal Assistance

There is general agreement that, if there is to be a U.S. SST program, the federal government must finance it, past flight test and possibly certification, by means of outright grants, advances to be recouped later, and/or guarantees of borrowings by the airframe and engine contractors from the financial community.

Three of the principal manufacturers and their suppliers lack both the ability and willingness to provide $4 billion or so for the development and production funds required before the date around 1975 at which time net cash inflows commence. Their collective net worths, in-

·cluding that of some of their probable subcontractors, do not approach such a figure. Besides, even if able, why should airframe and engine manufacturers risk such sums on an SST? They have alternative investment opportunities involving fewer technical uncertainties, smaller investments, and shorter "dry" periods before recovery of principal. That one airframe and one engine manufacturer are expected to be selected around January 1, 1967, for continued development with government assistance, almost guaranteeing a limited monopoly of a usable design eventually, is not enough apparently to evoke more than 10-20 percent participation in development costs by the manufacturers.

Potential airline customers are not expected to be a significant source of funds. No way remains of compelling the airlines to depart from their usual policy of "wait and see" before purchasing. Firm orders cannot be expected until after successful prototype flight tests in 1970 perhaps. And even then it seems unlikely that more than half the aircraft's flyaway price can be extracted through progress payments six to twelve months before delivery. (Although the first fifty or so SST's delivered probably have a special premium value of several million dollars each, because of high load factors during their initial two to three years of service, the first hundred-odd U.S. SST's have already been allocated for refundable deposits of $100,000 each.)

The combination of high technical risks and large dollar magnitudes makes it most improbable that the manufacturer, or indirectly the financial community, will provide the needed funds. Thus the federal government must either provide assurance against certain risks (e.g., accidental loss of a prototype) and/or become a sort of silent partner (providing say 80 percent of the net cash outflows and sharing say 80 percent of the net cash inflows). Such a scaling down of the private sums at risk to one-fifth of the total funds involved, with a commensurate scaling down of cash surpluses later, might be sufficient to evoke limited private financing despite the program's many uncertainties.

In addition to the funds required for development and production of an SST, safe and economical operation of SST's will be possible only if government expenditures are made for improved air traffic control, solar and cosmic radiation monitoring systems, and an improved meteorological forecasting capability.

The SST will cruise at 65,000 as compared to 35,000 feet for subsonics. The effects of wind and temperature on SST fuel consumption and the consequent impact on SST operating costs and payload make better data on these parameters vital. At supersonic speeds more advance warning is required to maneuver around unfavorable weather. In the period from 1970 through 1990 meteorological improvements alone could cost governments over a billion dollars.

These negative "externalities" should be considered part of the price of having an SST.

How Should the Government Recover?

The manufacturers will not risk even limited sums unless they and the government can agree on a financial plan that specifies at least the formula by which government will recoup its share of development and other "sunk" costs.

Unless there are significant and "external" national interests served by the program, a controversial issue discussed below, the federal government must be placed in a position to obtain the same rate of return on its investment as do the manufacturers if an impolitic and unwarranted subsidy is to be avoided. Thus if the program is continued because the $4 billions-odd worth of resources invested in it are expected to earn 10 to 15 percent before taxes or interest, and if the government were to accept say 5 percent on an 80 percent share of costs, the manufacturers would be expected to earn 30 percent to 55 percent on their investments. The only way government can avoid paying such a subsidy and still accept say 5 percent for itself, is to continue a program that prospectively wastes some of the nation's resources by promising a return less than that ordinarily obtained from domestic labor and capital used by industry in the U.S.

Thus government must be expected to share in profits if (1) the program appears economically justifiable and (2) unjustified subsidies to manufacturers are to be avoided. There is no logical escape. This requirement for government profit sharing has nothing to do with "socialism."

A major issue is whether government recovery should be either through some sort of tax or through sharing in net cash inflows as a silent partner.

Levies most often mentioned are:

1. A "royalty" or tax "off the top," added to the manufacturer's price of the aircraft: this would probably mean a $10 million tax added to say a $35 million manufacturer's price, and such a tax would lose sales to competing subsonics and possible supersonics.

2. An annual tax on SST aircraft operated by U.S. airlines: this would advantage foreign airlines, and a tax adequate for full government recovery plus an equity risk return would have the same present discounted value and incidence on price and sales as would an adequate royalty on delivery.

3. A tax of 10-20 percent on fares charged all SST passengers originating or terminating in U.S. territory: such a tax would further divert passengers to subsonic aircraft.

A tax on jet fuel purchased in the U.S. by all commercial superson-
ics may have to be reserved to pay for the special SST flight support
services discussed above.

The alternative to a tax is sharing through a financial partnership or
"pooling" arrangement. There could be one pool on the airframe and
another on the engine. In each case the selected manufacturer and the
federal government would be the initial members—able to sell their
shares later to financial intermediaries. They would contribute, quarter
by quarter, development and production costs in some agreed upon
ratio. And, when net cash inflows start around 1975, these would be
shared by the contractor and the government in proportion to their
credits in the pool. Such credits should include both advances and an
equity rate of return, preferably the same for each partner, of from 10
to 15 percent compounded.

One advantage of pooling as against taxing is that there is less likeli-
hood of handicapping U.S. SST manufacturers or U.S. airlines oper-
ating SST's.

Another potentially important advantage of pooling is that the self-
interests of the airframe and engine manufacturers are made coinci-
dent with the interests of the federal government. If the contractor be-
lieves some design change will more than pay for itself, i.e., that it will
earn a good rate of return on the extra cost of making a change, gov-
ernment will similarly be advantaged if the contractor knows his busi-
ness. And ordinarily, because airframe and engine manufacturers are
experienced in commercial air transport and presumably know the air-
lines' needs better than FAA officials, pooling could permit the federal
government to give the manufacturers the maximum degree of private
initiative possible. Such freedom may be needed to make the program
an economic success. Certainly, the SST program should not be admin-
istered as though the U.S. were developing and procuring a bomber
say, for hopefully it is airlines and not the federal government that
will buy and use this aircraft.

Another recoupment issue is whether the government should share
in manufacturers' receipts or net cash inflows forever or only until
such time as the government has realized some previously stipulated
rate of return on its outlays.

Making either the royalty or pooling claims to SST sales revenue
transferable would enable the government to sell its rights to future
revenues to the private sector if and when success of the program
seems assured.

Why should government seek to recover all its advances from those
who buy or use the SST, for are there not other broad national inter-
ests being served?

The most intangible and commonly suggested "external" benefits are (1) technological "fall-out," (2) contingent military use, and (3) national prestige, but each of these proclaimed grounds has been vigorously denied by others.

A more explicit argument is that the sale of U.S. SST's at $35 million or so each will benefit the U.S. balance of payments in the 1970's. Such claims ignore many substitution effects. Traditionally, half U.S. aircraft exports have been financed in the U.S. Each U.S. SST exported may mean at least one less U.S. subsonic exported. If the U.S. sells competitive aircraft to foreign airlines, U.S. airlines may lose passengers to them. Estimates suggest that, over five years from date of sale, the net balance-of-payments credit for the U.S. from the sale of an SST is no more than 5-10 percent of its U.S. export price (and may even be negative if increased U.S. tourist spending abroad is considered).

Nevertheless, some gross external benefits must exist, although not necessarily net of special high altitude meteorology costs, etc. In the end the program's Phase I and II (design competition) costs to the federal treasury might be "forgotten" as an alleged contribution of $300 million-odd to some vague "national interests." But beyond that, and starting with Phase III (prototype development) in 1967, the program should be treated as a commercial risk enterprise of manufacturers and government in partnership.

Mitigating the Monopoly

The federal government in early 1967 may select a single SST airframe and engine contractor to continue prototype development. The selected manufacturers will emerge with considerable monopoly power that could be of considerable value. How can the government use public funds to create a monopoly and yet protect the public interest without destroying the managerial perogatives and efficiency of private manufacturers and airlines?

First, the monopoly will be limited, for there may be some competition from the supersonic Anglo-French Concorde; but more important should be the rivalry of growth subsonics, competition that compels the selected U.S. SST manufacturers to provide the best possible performance at a price.

Second, the selected prime development contractors will have many subcontractors, especially on the airframe. These subcontractors will acquire a competence that later could be used by a rival prime contractor. Also key employees have been known to transfer at higher salary to a rival manufacturer.

Third, the federal government is expected to have the right to li-

cense patents at zero cost, together with shop drawings, to any other airframe and/or engine manufacturers who might later seek to compete.

Hence, if a really large and profitable SST market were to prove itself, it is not evident that rival manufacturers could not materialize. The DC-8 did follow the Boeing 707. A first generation manufacturer is not always sure of developing a profitable aircraft that excludes competition.

A single airframe and single engine manufacturer for the U.S. SST should create other novel problems more easily overlooked.

One is division of the flyaway aircraft's price between airframe and engine. There is a very large range of indeterminacy. The engine for example might be priced almost down to its marginal production cost. But its price could be as high as the aircraft price minus marginal airframe cost. The maximum conceivable engine price could be three times the minimum. Previously an airframe company had some choice among engine firms to narrow the range of possible engine prices.

The airlines will be face to face with a limited monopolist able to charge variable (i.e., discriminatory) prices unless prevented by government. Some of the earlier deliveries, assuming they can maintain schedules, have potentially higher load factors and hence should command premium prices. Should the competition of Concorde and the subsonics prove minor, some airlines may ask for federal ceilings on U.S. SST prices.

This could be a dangerous step. If government regulates SST prices, should it not more energetically regulate SST fares so that airlines do not receive unjustified profits, and perhaps depress load factors through excessive competition in flight frequencies on approved SST routes? The situation of the U.S. merchant marine should be enough to deter any airline management from inviting government intervention.

The remaining alternative would be to develop two airframe-engine combinations—which practically would mean also producing two rival aircraft. This would double development costs. It would also lose some "learning curve" efficiencies in manufacturing. These extra costs would be justified only if, attributable solely to prolonged competition, operating costs were reduced by about a quarter.

Are the Benefits General Enough?

Costly public enterprises of great inherent risk are more easily justified if the spending of taxpayers' funds occasions widespread benefits. It is not enough that they provide employment and profits for a few localities and firms. How diffused are the benefits of the SST program likely to be?

About 85 percent of U.S. residents have never flown, those who do

fly do not always take long-haul flights, and perhaps less than 5 percent of all Americans will ever fly SST's at their higher fares. Private, nonexpense account, long-haul passengers will mostly continue to fly subsonically. (It is not even certain that the federal government will reimburse its employees and those of its contractors for a supersonic surcharge.)

Further, American SST passengers will tend to travel to and from a few areas, such as New York, Chicago, Los Angeles, San Francisco, Seattle, Washington, D.C., and Miami. Americans living elsewhere may never use an SST except on international flights. But 100 million Americans may find themselves subjected daily to sonic booms if overland SST flights are permitted.

For all these reasons, and as the U.S. SST program is seen increasingly as a rival to expansion of Great Society programs, it is certain to become more controversial.

Conclusions

There are few modern instances of development with public funds of a technologically advanced product that is to be produced—very likely by a monopoly—and used commercially by private firms. Desirable guidelines in such cases are: (1) continuation only so long as the program is expected to earn a rate of return comparable to that expected by U.S. industry in making investments; (2) equal government sharing in such a rate of return, partly to avoid subsidization, but also to provide incentives for only economical design changes, etc.; (3) full recoupment by government, plus an equity or risk-taker's return, except insofar as their clearly are net "external" or national interests; (4) recoupment by government of its advances, not by taxes that reduce sales and use, but through some partnership (pooling) arrangement; (5) avoidance of special controls that will prevent the usual exercise of experienced management by the airframe and engine manufacturers; (6) use of public funds only to the extent that a private and excessively profitable monopoly will not be created; (7) no unrecovered government subsidies except where adequate benefits are likely to be diffused among a large fraction of citizens.

Finally, it is to be hoped and expected that a U.S. prototype of an SST will be flying well before the end of the 1970's, one that promises to be safe and profitable without being a public nuisance because of sonic boom. This means a state of art that will permit a rate of return of at least 10 percent on all resources invested in development after 1966 without supersonic flight over populated land areas. Until proposed designs can give this promise with more confidence it seems premature to begin construction of a prototype aircraft.

AN APPLICATION OF COST-BENEFIT ANALYSIS TO THE WORK-EXPERIENCE PROGRAM*

By WORTH BATEMAN

Department of Health, Education, and Welfare

Introduction

The Work-Experience and Training Program is one of many federally supported programs which aims to reduce poverty and dependency by raising individual capabilities for self-support. Funded under Title V of the Economic Opportunity Act of 1964 and administered by the Department of Health, Education, and Welfare, the program seeks to increase the employment and earnings potential of family heads who are recipients of public assistance (primarily the adult recipients in families receiving Aid to Families with Dependent Children, AFDC) and nonrecipients with similar characteristics.[1]

A variety of projects are funded in this program. Some are training projects utilizing the programs available under MDTA and the Vocational Education Act. Some provide adult basic education. Others are work-experience projects in which welfare clients and other needy persons are given jobs in public or nonprofit agencies. However, the majority of projects incorporate all these components.

The following profile shows some of the characteristics of the 101 thousand persons enrolled in the program in fiscal year 1966: participants are about equally divided between males and females; more than 60 percent of enrollees are married and over 90 percent are heads of households with four or more dependents; more than 80 percent are between 21 and 49 years old; nearly 30 percent are high school dropouts; over half have completed eight grades or less of formal schooling; assistance payments made under AFDC, allowances provided

[1] According to a recent Welfare Administration report, Title V differs from other employment programs "in that it concentrates on developing an intervention program for the entire family, not just for the employable adult . . . [These individuals] have individual or family adjustment problems as well as educational and vocational training deficiencies which deter or preclude them from employment." Although employment is the ultimate goal, "the more proximate objectives center around improved individual and family functioning." Although it may, in fact, be necessary to overcome family problems as a condition for achieving the employment objective, there is no mention of improved individual and family functioning (whatever that may be) in the legislation authorizing expenditures for this program. This emphasis, in part, may be the natural consequence of having the Welfare Administration administer the program. See Abraham S. Levine, "Cost-Benefit Analysis of the Work-Experience Program," *Welfare in Review*, Aug.-Sept., 1966, pp. 1-9.

under Title V of the Economic Opportunity Act, or General Assistance are the most important (and in most cases the only) sources of family income.

Thus, the program is focused on potentially employable poor persons with little formal education who are currently unemployed and who lack the means to support themselves and a comparatively large number of dependents. In comparison to the number currently enrolled in the program, this target population is large. Among public assistance recipients alone, it has been estimated that perhaps 250 to 350 thousand persons might be made self-supporting if given the proper training and other supporting services (e.g., day care for the children of AFDC mothers).[2] In addition, the universe of potential beneficiaries of the program includes many of the unemployed and employed poor who are not now recipients of public assistance.

The exclusive concern of this program with what Cain and Somers[3] have called the "specially disadvantaged" reflects a pronounced change in emphasis of federally supported training programs in the last several years. The programs of the War on Poverty, of which this is one, are the most well-known examples of this change. At present, very little is known about the effectiveness of these programs in raising the earnings of the groups they serve.[4]

The purpose of this paper is to: (1) develop an analytical framework for evaluating the work-experience component of the program; and (2) to estimate its potential effectiveness in improving the capability for self-support.

Measures of Benefit

The benefits of the Work-Experience Program can be divided into two parts. First, there are the short-run or immediate benefits of a work-relief program—in part, the output produced by people working who would otherwise be unemployed. Second, there are the long-run benefits of reduced dependency and improved potential for economic independence and self-support. These benefits can be estimated by comparing the present value of expected future earnings of program participants before and after the training or with those of individuals

---

[2] Robert H. Mugge, "Demographic Analysis and Public Assistance," prepared for presentation at the Annual Meeting of the Population Association of America, New York City, Apr. 30, 1966.

[3] Glen Cain and Gerald Somers, "Retraining the Disadvantaged Worker" (Sept., 1966, mimeo.).

[4] Cain and Somers report that on the basis of several case studies which they analyzed, the position of the disadvantaged worker is enhanced considerably when compared with their own pretraining experience and compared with disadvantaged workers who have not been trained. See Cain and Somers, op. cit., and Gerald Somers, "The Experience with Retraining and Relocation," Conference on Manpower Policy, Berkeley Unemployment Project, New York City, June 20-22, 1966.

with the same socioeconomic and demographic characteristics who have not participated in the program.[5]

Over any period of time, earnings will be equal to the product of the hourly wage rate which prevails for that period and the number of hours worked:

$$(1) \qquad\qquad E_t = H_t W_t$$

where E_t represents the earnings in period t, H_t is the number of hours worked during period t, and W_t is the prevailing hourly wage rate over period t.[6]

A change in earnings during period t can result from a change in the number of hours worked, from a change in the wage rate, or from a change in both:

$$(2) \qquad \Delta E_t = H_t \Delta W_t + W_t \Delta H_t + \Delta W_t \Delta H_t \text{[7]}$$

The first term on the right-hand side of (2) can be thought of as a "wage effect," i.e., an increase in the value of employee productivity for a given number of hours worked; the second term can be thought of as an "employment effect" which, for a given wage rate, reflects additional time worked; and the third term an interaction term or "combination effect" resulting from a simultaneous change in both factors.

Assuming all improvements in expected lifetime earnings are attributable to program participation, the increase in the present value of expected lifetime earnings can be represented by:

$$(3) \qquad \Delta E = \sum_{t=k}^{d} [S_t(H_t \Delta W_t + W_t \Delta H_t + \Delta W_t \Delta H_t)]...(1+r) \; \longrightarrow \; {}^{(t-k)}$$

[5] This, of course, is not equivalent to the conceptually superior method of comparing the earnings and employment experience of given individuals with and without the training program. The validity of the results depends on how good the surrogate data are for the with-without data. See Burton A. Weisbrod, "Conceptual Issues in Evaluating Training Programs" (Feb. 14, 1966, mimeo.).

[6] If the wage rate for part-time work differed from that for full-time work (1) would have to be reformulated as follows:

$$(1') \qquad\qquad E_t = (H_{ft} W_{ft} + H_{pt} W_{pt})$$

Where H_{ft} and H_{pt} are the number of hours worked full time and part time, respectively, and W_{ft} and W_{pt} are the wage rates for full- and part-time employment, respectively.

[7] Thus, a 10 percent increase in the number of hours worked has exactly the same impact on earnings as a 10 percent increase in wage rates. This alone has important policy implications since the same increase in earnings can be obtained through expenditures which increase the probability of employment or increase worker productivity. But the costs of doing so may be quite different. For example, in a simple model, we might have: $\Delta E = f(p)$, where p is the proportion of a given "human investment" budget spent on placement and referral activities (say like those of state employment offices) and $(1-p)$ is the proportion spent on productivity increasing activities (training programs, for example). Although we should maximize $f(p)$ for a given budget, it is not clear, a priori, that p will be less than or equal to one. Of course, the probability of employment is dependent not only on demand but also on willingness to work. The poor are little different in this regard than others. The higher the wage rate, *mutatis mutandis*, the higher the probability of employment. At very low wage rates, the supply curve may be very inelastic, and in the extreme case, it may be a discontinuous one. Below that point, the quantity of labor supplied would be zero.

fly do not always take long-haul flights, and perhaps less than 5 per-
cent of all Americans will ever fly SST's at their higher fares. Private,
nonexpense account, long-haul passengers will mostly continue to fly
subsonically. (It is not even certain that the federal government will
reimburse its employees and those of its contractors for a supersonic
surcharge.)

Further, American SST passengers will tend to travel to and from a
few areas, such as New York, Chicago, Los Angeles, San Francisco,
Seattle, Washington, D.C., and Miami. Americans living elsewhere
may never use an SST except on international flights. But 100 million
Americans may find themselves subjected daily to sonic booms if over-
land SST flights are permitted.

For all these reasons, and, as the U.S. SST program is seen increas-
ingly as a rival to expansion of Great Society programs, it is certain to
become more controversial.

Conclusions

There are few modern instances of development with public funds of
a technologically advanced product that is to be produced—very likely
by a monopoly—and used commercially by private firms. Desirable
guidelines in such cases are: (1) continuation only so long as the pro-
gram is expected to earn a rate of return comparable to that expected
by U.S. industry in making investments; (2) equal government shar-
ing in such a rate of return, partly to avoid subsidization, but also to
provide incentives for only economical design changes, etc.; (3) full
recoupment by government, plus an equity or risk-taker's return, ex-
cept insofar as their clearly are net "external" or national interests;
(4) recoupment by government of its advances, not by taxes that re-
duce sales and use, but through some partnership (pooling) arrange-
ment; (5) avoidance of special controls that will prevent the usual ex-
ercise of experienced management by the airframe and engine manu-
facturers; (6) use of public funds only to the extent that a private and
excessively profitable monopoly will not be created; (7) no unrecov-
ered government subsidies except where adequate benefits are likely to
be diffused among a large fraction of citizens.

Finally, it is to be hoped and expected that a U.S. prototype of an
SST will be flying well before the end of the 1970's, one that promises
to be safe and profitable without being a public nuisance because of
sonic boom. This means a state of art that will permit a rate of return
of at least 10 percent on all resources invested in development after
1966 without supersonic flight over populated land areas. Until pro-
posed designs can give this promise with more confidence it seems pre-
mature to begin construction of a prototype aircraft.

AN APPLICATION OF COST-BENEFIT ANALYSIS TO THE WORK-EXPERIENCE PROGRAM*

By WORTH BATEMAN

Department of Health, Education, and Welfare

Introduction

The Work-Experience and Training Program is one of many federally supported programs which aims to reduce poverty and dependency by raising individual capabilities for self-support. Funded under Title V of the Economic Opportunity Act of 1964 and administered by the Department of Health, Education, and Welfare, the program seeks to increase the employment and earnings potential of family heads who are recipients of public assistance (primarily the adult recipients in families receiving Aid to Families with Dependent Children, AFDC) and nonrecipients with similar characteristics.[1]

A variety of projects are funded in this program. Some are training projects utilizing the programs available under MDTA and the Vocational Education Act. Some provide adult basic education. Others are work-experience projects in which welfare clients and other needy persons are given jobs in public or nonprofit agencies. However, the majority of projects incorporate all these components.

The following profile shows some of the characteristics of the 101 thousand persons enrolled in the program in fiscal year 1966: participants are about equally divided between males and females; more than 60 percent of enrollees are married and over 90 percent are heads of households with four or more dependents; more than 80 percent are between 21 and 49 years old; nearly 30 percent are high school dropouts; over half have completed eight grades or less of formal schooling; assistance payments made under AFDC, allowances provided

[1] According to a recent Welfare Administration report, Title V differs from other employment programs "in that it concentrates on developing an intervention program for the entire family, not just for the employable adult . . . [These individuals] have individual or family adjustment problems as well as educational and vocational training deficiencies which deter or preclude them from employment." Although employment is the ultimate goal, "the more proximate objectives center around improved individual and family functioning." Although it may, in fact, be necessary to overcome family problems as a condition for achieving the employment objective, there is no mention of improved individual and family functioning (whatever that may be) in the legislation authorizing expenditures for this program. This emphasis, in part, may be the natural consequence of having the Welfare Administration administer the program. See Abraham S. Levine, "Cost-Benefit Analysis of the Work-Experience Program," *Welfare in Review,* Aug.-Sept., 1966, pp. 1-9.

under Title V of the Economic Opportunity Act, or General Assistance are the most important (and in most cases the only) sources of family income.

Thus, the program is focused on potentially employable poor persons with little formal education who are currently unemployed and who lack the means to support themselves and a comparatively large number of dependents. In comparison to the number currently enrolled in the program, this target population is large. Among public assistance recipients alone, it has been estimated that perhaps 250 to 350 thousand persons might be made self-supporting if given the proper training and other supporting services (e.g., day care for the children of AFDC mothers).[2] In addition, the universe of potential beneficiaries of the program includes many of the unemployed and employed poor who are not now recipients of public assistance.

The exclusive concern of this program with what Cain and Somers[3] have called the "specially disadvantaged" reflects a pronounced change in emphasis of federally supported training programs in the last several years. The programs of the War on Poverty, of which this is one, are the most well-known examples of this change. At present, very little is known about the effectiveness of these programs in raising the earnings of the groups they serve.[4]

The purpose of this paper is to: (1) develop an analytical framework for evaluating the work-experience component of the program; and (2) to estimate its potential effectiveness in improving the capability for self-support.

Measures of Benefit

The benefits of the Work-Experience Program can be divided into two parts. First, there are the short-run or immediate benefits of a work-relief program—in part, the output produced by people working who would otherwise be unemployed. Second, there are the long-run benefits of reduced dependency and improved potential for economic independence and self-support. These benefits can be estimated by comparing the present value of expected future earnings of program participants before and after the training or with those of individuals

[2] Robert H. Mugge, "Demographic Analysis and Public Assistance," prepared for presentation at the Annual Meeting of the Population Association of America, New York City, Apr. 30, 1966.

[3] Glen Cain and Gerald Somers, "Retraining the Disadvantaged Worker" (Sept., 1966, mimeo.).

[4] Cain and Somers report that on the basis of several case studies which they analyzed, the position of the disadvantaged worker is enhanced considerably when compared with their own pretraining experience and compared with disadvantaged workers who have not been trained. See Cain and Somers, op. cit., and Gerald Somers, "The Experience with Retraining and Relocation," Conference on Manpower Policy, Berkeley Unemployment Project, New York City, June 20-22, 1966.

with the same socioeconomic and demographic characteristics who have not participated in the program.[5]

Over any period of time, earnings will be equal to the product of the hourly wage rate which prevails for that period and the number of hours worked:

$$(1) \qquad\qquad\qquad E_t = H_t W_t$$

where E_t represents the earnings in period t, H_t is the number of hours worked during period t, and W_t is the prevailing hourly wage rate over period t.[6]

A change in earnings during period t can result from a change in the number of hours worked, from a change in the wage rate, or from a change in both:

$$(2) \qquad\quad \Delta E_t = H_t \Delta W_t + W_t \Delta H_t + \Delta W_t \Delta H_t[7]$$

The first term on the right-hand side of (2) can be thought of as a "wage effect," i.e., an increase in the value of employee productivity for a given number of hours worked; the second term can be thought of as an "employment effect" which, for a given wage rate, reflects additional time worked; and the third term an interaction term or "combination effect" resulting from a simultaneous change in both factors.

Assuming all improvements in expected lifetime earnings are attributable to program participation, the increase in the present value of expected lifetime earnings can be represented by:

$$(3) \qquad \Delta E = \sum_{t=k}^{d} \left[S_t (H_t \Delta W_t + W_t \Delta H_t + \Delta W_t \Delta H_t) \right]...(1+r)^{-(t-k)}$$

[5] This, of course, is not equivalent to the conceptually superior method of comparing the earnings and employment experience of given individuals with and without the training program. The validity of the results depends on how good the surrogate data are for the with-without data. See Burton A. Weisbrod, "Conceptual Issues in Evaluating Training Programs" (Feb. 14, 1966, mimeo.).

[6] If the wage rate for part-time work differed from that for full-time work (1) would have to be reformulated as follows:

$$(1') \qquad\qquad\qquad E_t = (H_{ft} W_{ft} + H_{pt} W_{pt})$$

Where H_{ft} and H_{pt} are the number of hours worked full time and part time, respectively, and W_{ft} and W_{pt} are the wage rates for full- and part-time employment, respectively.

[7] Thus, a 10 percent increase in the number of hours worked has exactly the same impact on earnings as a 10 percent increase in wage rates. This alone has important policy implications since the same increase in earnings can be obtained through expenditures which increase the probability of employment or increase worker productivity. But the costs of doing so may be quite different. For example, in a simple model, we might have: $\Delta E = f(p)$, where p is the proportion of a given "human investment" budget spent on placement and referral activities (say like those of state employment offices) and $(1-p)$ is the proportion spent on productivity increasing activities (training programs, for example). Although we should maximize $f(p)$ for a given budget, it is not clear, a priori, that p will be less than or equal to one. Of course, the probability of employment is dependent not only on demand but also on willingness to work. The poor are little different in this regard than others. The higher the wage rate, *mutatis mutandis*, the higher the probability of employment. At very low wage rates, the supply curve may be very inelastic, and in the extreme case, it may be a discontinuous one. Below that point, the quantity of labor supplied would be zero.

where S_t is the probability that individuals in year t will survive at least to year $t + 1$; r is the discount rate; and d is the year when earnings are assumed to cease.

In the short run, the benefits of the program can be measured by the value of the additional output produced by those individuals in just the work-experience component of the program. Assuming a competitively determined wage rate and no external benefits, this would be just equal to the earned assistance payments since an individual works that number of hours which, when multiplied by the wage rate, equals the public assistance payment for which he is eligible. Of course, these wage rates are not determined entirely by competitive forces since, for the most part, individuals are placed on government payrolls. Moreover, even assuming competitively determined wage rates for regular government employees, it would not follow that the value of the marginal product of a participant necessarily equals the wage rate. If that were true, it would seem that employment could be obtained at that wage and the necessity for being placed in such a job under the auspices of the program would disappear.[8]

What is really required is an estimate of the social value of the marginal product (i.e., the aggregate amount which individuals would be willing to pay for this output if they revealed their preferences). This may diverge from the wage cost. For example, having the grass neatly cut and trimmed along the side of the road by work-experience enrollees may cost X. A well-manicured parkway is a public good for which individuals using it may be willing together to pay more or less.

This may be represented by:

(4) $$SVMP_t = \alpha\beta R_t H_t,$$

where $SVMP$ is the social value of the marginal product; α is the ratio of the value of the marginal product of a work-experience enrollee to that of nonenrollees who could be used to perform the work; β is a factor reflecting the divergence of the $SVMP$ and the wage cost; R is the hourly wage rate; and H is the number of hours worked while enrolled in the project. If $\alpha < 1$, the $SVMP$ is reduced by using a work-experience participant to do the job rather than a more productive worker.

The sum of (3) and (4) is an estimate of the economic benefits of the program.[9]

[8] Except in those cases where governmental budget constraints prevent all activities being supported at a level where the social value of the marginal product equals marginal social cost; e.g., "public goods." See Paul Samuelson, "The Pure Theory of Public Expenditure," *Rev. of Econ. and Statis.*, Nov., 1954, pp. 387-89.

[9] Both the short- and long-run measures of benefit described above represent real returns to expenditures on the Work-Experience Program. There is a third category of real benefits accruing to the economy which might arise out of the income transfers made in this pro-

Unfortunately, the complete lack of data prevented the benefit analysis from proceeding further than a mere formulation of how it should be done. The earnings profile of individuals enrolled in this program is unknown and the follow-up data on post-enrollment earnings is completely unreliable. It is also not possible to estimate the critical parameters in (4) from existing data.

The strategy adopted at this point was to estimate how much the expected income of participants would have to rise just to "break even." This requires an estimate of what the present value of future earnings of participants would have been in the absence of the program and an estimate of program costs. Although the latter are available, the earnings data of participants are not and have to be constructed.

There are a number of problems associated with making such estimates. The 1960 Census provides the most complete information on earnings, income, and labor force status by age, sex, color, and educational attainment. But the published reports of the 1960 Census also have gaps in information and deficiencies in the form reported critical to this analysis. The most serious of these is the absence of earnings data as distinct from income data for females.

Earnings data are clearly the most appropriate for this analysis. The income data in the 1960 Census include public assistance, social security, unemployment insurance, etc., which tend to narrow earnings differentials among various socioeconomic groups. This means that the use of income data would tend to bias the break-even estimates downward. For males, however, data are available from published sources in the 1960 Census on earnings, classified according to the selected control variables. Since the Census data are grouped by age class, average

---

gram (or put differently, from the payments made to enrollees for work performed on work-experience projects). Such income transfers may have an important investment component in the sense that the conditions in which children grow up have an important bearing on their income earning capacity as adults. Income transfers to poor families are, in part, investments in a better home environment, diet, and health, which in turn, enhance the long-range economic prospects of poor children. There are, of course, other benefits which might be realized from the program by participants and their families. These might include increased satisfaction, *mutatis mutandis*, of working rather than being idle, the benefits to children of seeing the example of a working parent, and the increase in family stability. (The program permits public assistance payments to be made to families where the male is present irrespective of whether a state has adopted the unemployed parent component of the AFDC Program. Unemployed or marginally employed heads of families not enrolled in the Work-Experience Program and residing in states without AFDC-UP may find that desertion is the only option available for supporting or improving the support of their families.) The most significant external benefits which have not been accounted for in (3) and (4) are the reduced welfare payments and increased tax revenues which result from an increased capability for self-support. Although these distributive effects are not additive to real returns, they are the consequence of a successful project and should be considered in Program evaluation. See Burton A. Weisbrod, "Preventing High School Dropouts," in Robert Dorfman, ed., *Measuring Benefits of Government Investments* (Brookings Institution, 1963), pp. 136-39; and also, "Income Redistribution Consequences of Government Expenditure Programs," Second Conference on Government Expenditures, Brookings Institution, Sept. 15-16, 1966.

earnings by individual year of age were computed by imputing the average earnings of an age class to the mid-point of the class and then making linear interpolations between mid-points. This was done for white and nonwhite males in each education category. Earnings were arbitrarily assumed to stop at age 64.

In computing expected earnings in each age, education, and race category, an adjustment was made for the probability that an individual of a given age will survive to the next year by applying life table values by age, color, sex (no educational attainment) to the estimated earnings for each year of age.

Estimates of the expected probability of employment were obtained from the 1960 Census. At the present, it has not been possible to locate more recent information on labor force participation and unemployment rates classified according to the selected control variables. The effect of this procedure in the absence of any further adjustments is to bias the break-even points upward since overall employment conditions were less favorable in 1959 (the year for which earnings are reported in the 1960 Census) than currently.[10]

Measures of Costs

The estimate of costs made for this analysis is the additional cost for an individual participant which is incurred by the government (state and federal). While in general these are primarily additional administrative costs, work-related expenses, and public assistance payments, the actual estimate of these incremental costs varies by state,

---

[10] The 1959 male unemployment rate was 5.3 percent compared to 4.0 percent in 1965. For white males the rates were 4.6 percent and 3.6 percent and for nonwhite males 11.5 percent and 7.6 percent in 1959 and 1965, respectively. It should be pointed out that the estimate of average annual earnings which is desired for each age, education, and sex category is given by:

$$(6') \qquad \left[\sum_i^n (WH)_i \right] (n + p + k)^{-1}$$

where W is the hourly wage rate, H is the number of hours worked during the year, n is the average number of individuals in the labor force who had earnings during the year, p is the average number of individuals in the labor force during the year who had no earnings, and k is the average number of individuals who were not in the labor force. What we have estimated, on the other hand, is

$$(7') \qquad \frac{\left[\sum_i^n (WH)_i \right]}{n} \frac{(n' + p')}{n' + p' + k'} \frac{n'}{n' + p'} = \frac{\left[\sum_i^n (WH)_i \right]}{n} \frac{n'}{n' + p' + k'}$$

where all the terms in the expression are defined as above with the prime signifying that the observation is for a point in time rather than for a full year. The first term on the left-hand side of the equation is the average annual earnings reported in the Census. The second term is the labor force participation rates and the third term is the unemployment rate, the latter two observations being made at a point in time. Over the course of a year, $n' + p' + k'$ will, for all practical purposes, equal $n + p + k$. If $n' = n$, then (6') and (7') are equivalent. If n' is observed at a time of less than average unemployment, then $n' > n$. Of course, the converse is also true.

since these costs depend on what payments the individuals would have received from the government in the absence of the Work-Experience Program. This in turn depends on whether a state meets or falls short of meeting 100 percent of "need"[11] for public assistance recipients and whether the state has or does not have the Unemployed Parent (UP) component of the AFDC Program.[12]

The table below shows the components of incremental costs according to the category into which a state falls.

COMPONENTS OF INCREMENTAL COSTS OF WORK-EXPERIENCE PROGRAM

	States Which Meet Less Than 100 Percent of Need	States Which Meet or Exceed 100 Percent of Need
I. States with AFDC-UP and AFDC	ΔC_{11} = Additional administrative costs+work-related expenses+additional public assistance payments necessary to meet 100 percent of need	ΔC_{12} = Additional administrative costs+work-related expenses

II. States with AFDC but no AFDC-UP
  a. AFDC
  b. Male W-E participants
	ΔC_{21} = Same as above
	ΔC_{31} = Additional administrative costs+work-related expenses+100 percent of public assistance payments
		ΔC_{22} = Same as above
		ΔC_{32} = Additional administrative costs+work-related expenses+100 percent of public assistance payments

The additional administrative costs associated with program operations are self-explanatory and are a component of the incremental costs for all states. Work-related expenses such as transportation, clothing, etc., incurred only for program participants are also part of the marginal costs for all states.

Title V requires that work-experience participants receive payments equal to 100 percent of what the state defines as need for that individual and his family. In those states meeting 100 percent of need there are no additional public assistance costs incurred for welfare recipients participating in a work-experience project. However, public assistance recipients in states meeting less than 100 percent of need can increase

[11] Each state determines a level of "need" for public assistance recipients. The level of need is based on the prices of a market basket of goods and services deemed to be "adequate." The level established varies by family size and composition. Moreover, there is wide state by state variation in defined need and not all states allocate sufficient resources to public assistance programs to meet 100 percent of need. For example, Indiana defines need for an AFDC family of four to be about $225 per month whereas for the same size family, New York defines need to be about $255 per month. Although New York meets 100 percent of need, actual monthly payments in Indiana are about $110 or 48 percent of need.

[12] The 1962 amendments to the Social Security Act permitted federal matching of state expenditures for Aid to Families with Dependent Children where the male is unemployed. To date, only twenty-one states have adopted this provision of the law.

their income by enrolling in a project. The federal government pays the difference between the state grant and state defined need.

In those states without AFDC-UP, unemployed male heads of families are not eligible for federally aided public assistance. Thus, for a male enrollee in a work-experience project in such a state, the incremental costs include the full public assistance grant for which he and his family become eligible.

Other things equal,[13] the relationship among the incremental costs shown in the table above is as follows:

(5) $\Delta C_{32} = \Delta C_{31} > \Delta C_{21} = \Delta C_{11} > \Delta C_{22} = \Delta C_{12}$

and in the absence of differentials in benefits the benefit-cost ratios would rank from lowest to highest in the same order as presented in (5).

The definition of program costs presented above raises a thorny problem when one recognizes the income redistribution objectives of this program which are held by many of its supporters. The costs per participant as defined above are viewed as an investment in present and future productivity. In a sense, this does involve implicitly a distribution objective with respect to income. It is possible, however, to go further than this and to insist that the program is merely a way of making a number of needy individuals eligible for public assistance who would not be otherwise because of existing or potential public hostility to the AFDC-UP Program. If that were true, one could argue that the straight public assistance component should not be counted as part of the incremental costs of the Program since income redistribution now is what is being bought and not exclusively improvements in productivity in the future. The redistributional objective is in a sense accomplished with the help of a ruse and a little work. If the program does raise future income and contribute something to current GNP at the same time, then that is all well and good; but the definition of the incremental costs of doing this, one could argue, should be the same as in those states with AFDC-UP meeting at least 100 percent of need (i.e., comprising only administrative costs and work-related expenses).

The definition of costs given above and the benefit-cost analysis made below would also be critically affected by many of the major income maintenance proposals currently under discussion. For example, the adoption of federal minimum standards for the current categorical public assistance programs and the compulsory adoption by states of the AFDC-UP Program[14] would reduce interstate differentials in pub-

---

[13] That is, assuming no state-by-state variations in average need and payments per case, work-related expenses, and administrative costs.

[14] Both of these measures were recommended in the *Report of the Advisory Council on Public Welfare* (June, 1966).

lic assistance costs of the Work-Experience Program. The same, of
course, would be true if categorical public assistance programs were
replaced entirely by a guaranteed annual income or negative income
tax.[15]

Because of the dual nature of the objectives of this program, incre-
mental costs were calculated: (1) including incremental public assis-
tance costs and (2) excluding all public assistance costs. Costs were
estimated for male work-experience participants in Pennsylvania,
Michigan, Minnesota, and Kentucky.

"Break-Even" Analysis

The ratio of the incremental costs and the estimated present value
of future earnings is an indication of how much earnings would have to
rise for the program to break even. The results for white males ages
25, 35, 45, and 55 for one state, Kentucky,[16] are shown in Table 1.

The implication of this table is that, even when incremental public
assistance costs are counted, only a small percentage increase in future
earnings would be necessary to make the program break even. If work-
experience can increase the present value of future earnings of partici-
pants by as much as 2 percent at younger ages or 3 percent at upper
ages, it will pay for itself. If incremental public assistance costs are ·
not counted as costs of work experience, then even small increases
(less than .5 percent) would make the program break even.[17] However,
in a world of constrained budgets, breaking even is not a sufficient con-
dition for funding a project. If the desirability of funding this Program
is to be weighed against other alternatives for which benefit-cost ratios
have been estimated, it is necessary to make some guesses about the
actual effectiveness of the program in raising future earnings.

We can get a somewhat better "feel" for the problem by considering
that an individual earning $1.50 per hour, working forty hours per
week, fifty weeks in the year, earns $3,000 per year. If we assume that
an individual without work experience would continue to earn that
sum for forty years, the present value (discounted at 4 percent) of his
future earnings is about $59,000 (or roughly that of a white male 25
years old with 0-7 years of educational attainment; see Table 1, line
1). An increase of 25 cents an hour (i.e., a 17 percent increase) in his

[15] For a discussion of these income maintenance programs, see the *Report of the Ad-
visory Council on Public Welfare* and two recent articles in *The Public Interest:* James
Tobin, "The Case for an Income Guarantee," Summer, 1966; and Alvin L. Schorr, "Against
a Negative Income Tax," No. 5, Fall, 1966.

[16] Because of the similarity of the results the break-even points for the other three states
are not shown.

[17] It should be kept in mind that the earnings of work-experience enrollees are likely to
be below the average which we have estimated from the Census for each age and education
category. This biases the break-even points downward.

TABLE 1

PROGRAM COSTS, EARNINGS, AND BREAK-EVEN POINTS FOR WHITE MALE
PARTICIPANTS IN THE WORK-EXPERIENCE PROGRAM—KENTUCKY

Age/Education of Participants		Program Cost per Participant Column (1)*	Estimated Future Earnings per Participant without Program Column (2)†	Percent Increase in Future Earnings Required for Program to Break Even (1) as Percent of (2)
Age 25				
Elementary	0–7..	$1,064	$56,577	1.9%
	8..	1,064	75,938	1.4
High school	1–3..	1,064	88,577	1.2
	4..	1,064	99,652	1.1
Age 35				
Elementary	0–7..	1,064	52,719	2.0
	8..	1,064	69,555	1.5
High school	1–3..	1,064	81,378	1.3
	4..	1,064	91,596	1.2
Age 45				
Elementary	0–7..	851	40,532	2.1
	8..	1,064	53,041	2.0
High school	1–3..	1,489	61,569	2.4
	4..	851	68,977	1.2
Age 55				
Elementary	0–7..	638	22,325	2.9
	8..	851	29,149	2.9
High school	1–3..	851	33,789	2.5
	4..	1,064	37,694	2.8

* Includes public assistance costs.
† Computed from data in U.S. Bureau of the Census, *1960 Census of Population*, PC(2)-7B, Table 1 and PC(2)-5B, Table 4; and U.S. Public Health Service, National Vital Statistics Division, *Vital Statistics of the United States, 1961*, Vol. II-Sec. 2, *Life Tables*, Table 2–3. Future earnings were discounted at 4 percent.

expected wage rate would imply under the assumption above an increase in the present value of his future earnings of about $10,000 or a benefit-cost ratio of 10:1.[18]

Although these data are suggestive of large potential payoff, the fundamental fact remains that such calculations are pure guesswork and the fragmentary program information which is available permits estimates which are only slightly superior.

A three-month follow-up survey of work-experience participants who had terminated their enrollment showed that the average monthly wage of those employed was about $250. But only 42 percent of those located whose records were reported correctly were employed; 58 percent were unemployed.[19] It is possible to make a crude estimate of

[18] This, of course, does not include an estimate of the social value of the marginal product while enrolled in the project.
[19] Of the total 14,221 terminees in the sample, only 8,370, or 59 percent, were located or had records which were reported correctly.

what the rate of unemployment would have been in the absence of the program from data recently released by the Labor Department.[20] Public assistance recipients who had received training under the MDTA through February, 1965, had been unemployed 54 percent of the year prior to enrollment for training.[21] Assuming this is representative of the fraction who would be unemployed at any given time during the year, given the crudeness of the comparison, and the other factors (e.g., changes in aggregate demand for labor) which have not been accounted for, there would appear to be little difference between the employment record of Program participants before and after work experience. The post-training earnings of the employed MDTA trainees averaged about $250 a month, or the same as that of work-experience enrollees.

The most significant differences between the public assistance recipients trained under MDTA and the work-experience participants was the much larger proportion of the former who were employed after training: 64 percent compared to 42 percent. This, however, is not surprising. The provision of work opportunities for individuals likely to suffer long periods of unemployment may be effective in preventing the loss of working skills demanded by almost all jobs. However, it seems doubtful that any significant increases in earning capabilities will be realized unless there are real opportunities for training and retraining. At present, however, the emphasis on manpower training in the Work-Experience Program is not great. In general, individuals are placed first in jobs available within state and local governmental units which, in many cases, do not provide any significant upgrading in the skills of the participants. The program could most likely be made more effective by infusing it with a much larger training component. This could be accomplished by expanding the MDTA programs of institutional and on-the-job training which are directed specifically to the disadvantaged. Of course, the costs would be substantially greater than the costs of work-relief.

[20] U.S. Department of Labor, "Training of Public Assistance Recipients Under the MDTA," *Manpower Evaluation Report No. 6*, Apr., 1966.

[21] This is admittedly a very rough means of obtaining the needed control data. However, the characteristics of the MDTA public assistance trainees are not too dissimilar from those of the work-experience enrollees: they were about equally divided between males and females; they had approximately the same age distribution; they were preponderately the heads of families. The MDTA trainees, however, had a higher average level of educational attainment than did the work-experience enrollees.

A PROPOSED METHODOLOGY FOR COMPARING FEDERALLY ASSISTED HOUSING PROGRAMS*

By WILLIAM B. ROSS

Department of Housing and Urban Development

Sheerly by coincidence, the launching of the new Federal Department of Housing and Urban Development followed shortly on the heels of the new presidential directive[1] for increased formal emphasis on more precise identification of national goals and on more systematic analysis of alternative means of reaching them most effectively. From the beginning, there have been no illusions either that the desired objective of effective public administration was a new one or that all the sought-after improvement in the precision and quality of analysis would be achieved overnight—or in a single year—or in a single administration.

The urgency of the presidential instruction was clearly of the "let us begin" variety. But the full burden falls on individual agencies to translate into their own operational setting an analytic approach long pioneered—but still conceded to be in the development stage[2]—in the analysis of defense goals and alternatives. The launching of so many simultaneous diverse efforts—without benefit of the extensive developmental work which preceded the formal Defense Department system—has inevitably resulted in analysis ranging from the uneven to the unsatisfactory and, often, to the unproduced. Each agency making this effort faces its own particular version of generalized national goals supported by broad individual program objectives leading to loosely related program activities on which voluminous records are kept on everything but how the program activities support the program objectives and lead specifically and measurably to the fulfillment of concrete national goals.

When everything depends on everything else, it is not immediately clear whether better analysis can be started most effectively: at better specification of national goals; at clearer development of relationships

between overall program objectives and national goals; at testing of
the relationships between specific program activities and program ob-
jectives; or at development of better data on any or all of the preceding
relationships?

The present paper is, at best, a progress report on how one analytic
staff in one agency is attempting to approach one specific policy prob-
lem cycle. The aim of the paper is to describe what we believe to be
useful analytic methodology for the federal decision-maker; we try to
be explicit about the stages of analysis in which we choose to defer
concentration while presenting for critical review those tentative pro-
posals which now appear to us both to be relevant and to lead in useful
directions.

Federal Goals in Housing

Since 1949, the nation has had the expressed goal "of a decent home
and a suitable living environment for every American family. . . ."[3]
The directness, simplicity, grandeur, and scope of this expression have
been the envy of presidential speechwriters and of legislative drafts-
men ever since. None has topped it, and most have sensed the futility
of even trying.

But the process of refining this broad goal into operational terms—
even setting aside for present purposes the phrase "a suitable living
environment"—is neither direct nor unambiguous. Certainly, the
refinement process must consider the broad variety of action programs
enacted under the "decent home" banner. Not all of the relevant pro-
grams are the responsibility of the Department of Housing and Urban
Development, but even those that are represent a diversity which we
feel incapable of encompassing altogether within a single analytic
framework.

For initial clarity, we have chosen somewhat arbitrarily to subdivide
HUD's housing programs between those which appear to fit most logi-
cally under the phrase "to facilitate effective private housing market
mechanisms" as distinguished from programs whose ultimate purposes
appear to include overt assistance in the provision of housing at below-
market costs. This dichotomy is neither a clean one nor of indefinite
value. But, for the present, we find it to be a useful distinction.

The "assisted" housing programs are not themselves by any means a
homogeneous group. Interpretation and interpolation of recorded legis-
lative histories, inductive reasoning from obvious relationships, and
even interrogation of participants in the legislative process all have a
part in trying to identify just what unique combination of contributions a

---

[3] Declaration of National Housing Policy, Section 2, Housing Act of 1949 (Public Law
171, 81st Cong.; 63 Stat. 413).

particular program was (and is) expected to make toward the overall decent home goal.

Essentially, each well-established federal activity has been the product of a unique, continued exercise in group dynamics involving interaction between (1) proposers of legislation (executive and legislative); (2) subcommittees, full committees, and leadership groups in the legislative process; (3) executive branch program administrators; (4) executive and legislative participants in the budgetary process; and (5) the "target group" of individuals or institutions. The net programmatic result has both the strengths and the weaknesses of the processes of democracy; above all, however, it means that single-dimensional measurement and single-criteria evaluation and even single-disciplinary study are inadequate—or even misleading.

The unraveling of the mixture of motivations and goals is the most difficult and the most critical part of analytic efforts—and, potentially, the most. productive. But, the complexity of these patterns means that analysis cannot start at a logical beginning of clear identification of unambiguous goals and proceed in orderly, martial fashion to conclusions. The program grouping with which this paper deals is merely one special case of this problem.

At this stage, for example, we do not find it useful to insist on a direct separation of the income-redistribution goals of housing assistance programs from the developmental goals—both economic and social— of providing more housing of socially acceptable quality for various population groups. Yet, we cannot ignore the existence of "low" income groups too "poor" to be aided directly through existing housing assistance mechanisms. The evidence of constraints on the utility of functional programs for income-maintenance objectives requires more careful delineation of objectives/means relationships but cannot be regarded as a prima facie case for abandonment.

The Concept of Output

In the past, it would have been relevant, or even necessary, to debate at length whether the "outputs" of housing assistance programs should be quantified in "goods" terms (e.g., units produced) or "services" terms (e.g., unit months of occupancy). This dilemma appears to us to have been resolved by increased reliance on upgrading of existing substandard housing units and by the initiation of new assistance methods which involve temporary support of units for varying periods of time. Thus, for the present, at least, we find it necessary to use output concepts of the flow-of-service in order to compare those programs which we now administer, let alone the infinite varieties of alternative assistance devices which could be considered.

Again, the choice is a tentative one, and the probabilities are high that we will, for some parts of the analysis, need to return to production or "goods" output measurements. For either choice, considerably more thought will need to be given to the "standard-substandard" concepts and definitions.

"Housing Assistance" Programs

Any listing of federal "programs" is tentative and subject to expansion or contraction as analysis focuses, at one stage, on relevant characteristics which require separate consideration of distinctions within an administratively cohesive program or, at another stage, on characteristics which blur the distinctions between programs administered by totally different organizations or agencies.

With these reservations, we consider that seven administratively separate programs constitute the "assisted housing" category:

1. Low-rent public housing (including new construction; acquisition of new units; acquisition of existing units, with or without rehabilitation; and leasing of existing units).

2. Rent supplements with mortgage insurance (with or without federal acquisition of the insured mortgage).

3. Direct loans at below-market interest rates for rental housing for the elderly and handicapped.

4. Below-market interest rate loans to higher educational institutions for housing and related facilities (including faculty housing and married student apartments).

5. Insured mortgages at below-market interest rates coupled with federal acquisition of the insured mortgage.

6. Below-market interest rate loans for rehabilitation of owner-occupied dwellings in urban renewal areas.

7. Partial grants to low-income homeowners in urban renewal areas for rehabilitation.

Economic Characteristics of "Assisted Housing" Programs

While this group of programs has the common economic characteristic that the housing for some period of time is provided to the occupant at less than the private market costs, the significant differences within the group may be categorized, tentatively, as: (1) differences in degree (level or proportion) of assistance; (2) differences in time pattern of gross and net federal outlay or cost; (3) differences in time pattern of benefits; (4) differences in the physical unit provided; (5) differences in total costs of the unit provided.

For purposes of this paper, it is necessary only to describe the range of these differences rather than the full array of them.

The degree of assistance can be as high as a partial contribution to operating costs plus the full capital costs of building new units as represented by the maximum annual federal contribution under low-rent public housing; it can be as low as a small fraction of the cost of a fix-up job as in the owner-occupant rehabilitation assistance programs. It may be as fixed as the maximum dollar limit for rehabilitation grants; it may be as indeterminate as the rent supplement payment which depends more on the future incomes of the tenants than it does on initial construction costs.

The time pattern of net federal outlays or cost can be as simple and as short as the one-check disbursement of a rehabilitation grant; it can be as long as the fifty-year maximum repayment period on a college dormitory loan; it can be as complex as the forty years of annually calculated net federal contributions to a low-rent public housing project plus the estimated federal loss on tax-exemption of the bonds sold to finance the project.

The time pattern of benefits can last as long as the housing unit stands in many cases; it could change or even terminate at the first annual tenant income reexamination of a rent supplement project; it can be linked to specific time periods as in the case of public housing leased units; or a residual part of it could be capitalized into the net selling price of a unit rehabilitated with grant assistance when the property subsequently changes hands.

The physical units provided show large distinctions within each program group as well as between them—even when reduced from "structures" to "dwelling unit" terms. The range can be from the sleeping/study room shared by two or four college students to the multibed-room apartment for a large family in public housing or a rent supplement project.

Over and above the unit cost differences attributable to differences in average unit size, there are other major cost factors of great policy relevance. Quality of construction standards are among the most significant of these. Quality standards of size, equipment, and allowable amenities are administratively established pursuant to general legislative standards or guidance (e.g., "not be of elaborate or extravagant design or materials")[4] and thus differ most between programs which have been administered separately. The scope of "related facilities" included can also be significant as when a public housing project includes community rooms for recreational or social services activities. Equal size and quality units can be provided in large- or small-scale projects or in single units with greatly differing production costs. And,

[4] Section 401(a), Housing Act of 1950 (Public Law 475, 81st Cong.; 64 Stat. 48).

highly significantly, local land values and construction costs for the same product can differ within the nation by nearly 100 percent.

Relevant Policy Issues

Given a going concern which carries on, among its activities, a series of related programs with this range of diversity, the policy questions most frequently raised are likely to be along the lines of: What changes, if any, ought to be made in the multitude of administratively established constraints? What alternative devices could be proposed to better serve all or part of the objectives of this program group? And, most of all, what should be the incremental change in the relative proportions in which these related—and sometimes complementary—programs should be offered in the next annual budget?

The first of these questions is literally always relevant. The others become pressing on at least an annual cycle and become increasingly intense in those years in which a program exhausts its previously authorized time or money limits and must, in effect, have its option renewed. In each such policy question, varying degrees of partial goal reexamination and refinement are necessarily involved, but the complex nature of the composite goal mixture seldom permits abstract, total goals formulation except in the context of a policy issue along one of these lines.

At this stage of development, our approach to the assisted housing program group focuses on the income redistribution (housing cost savings to tenants) aspects which are their most common denominator but not necessarily the dominant goal of any one program nor of equal weights in the goal mix as between programs.

Measuring "Costs"

For present purposes, we believe it appropriate to measure "costs" of the assisted housing programs in terms of net federal payments, although more sophisticated measures may become more useful at later stages of analysis; e.g., in dealing with macroeconomic aspects of national housing investments. Our federal cost measure cannot, however, be restricted solely to direct outlays and recoveries of the Department. The nature of financing mechanisms used in this program group requires consideration of interest costs (only parts of which may be reflected in the agency's books) and the effects of exemption from federal income taxes. These factors are considered not for the sheer joy of complicating the analysis but solely out of the necessity that they be taken into account in reducing the disparate programs to a meaningful common basis.

The consideration of interest costs is present in another form when

we attempt to make comparisons between the disparate time patterns of disbursement and repayment possible within this program grouping. Discounted present values are an obvious device for handling this problem, and the use of federal borrowing rates is an equally obvious first approximation of an appropriate rate of discount when we are comparing alternative federal methods of spending money to achieve the same or related ends. Again, more complex discount rate concepts, e.g., taxpayer's marginal value of consumption foregone, are only appropriate in the analysis of more aggregative aspects of the housing problem.

Measuring "Benefits"

Isolating the income-redistributional results of the housing assistance programs can be regarded as a first approximation of one aspect of the "benefits" of these programs, although we will most certainly want subsequently to look at the pattern of "savings" against the beneficiary income patterns to make a judgment on how well the income-redistributional results correspond to public goals. Initially, we may assume that the assisted tenants are those for whom income benefits are intended to accrue.

While the rents (or monthly housing costs for owner-occupants) of the units aided under the various programs are determinable, we need a base concept against which to compare these rents. Comparison with rents paid by the tenants in their previous housing has immediate but superficial attractiveness. It suffers from the defect that varying quality, size, and location of the previous unit make this comparison unreliable at best.

A better concept—which both avoids the uncertainty-of-previous-condition problem and adapts to the variety of factors influencing unit costs—is that of "private market housing of comparable quality." This standard has considerable flexibility in analysis and appears to be useful for a wide variety of comparisons, provided one keeps in mind the weakness of its implicit assumption that the tenant would have chosen housing of this size, quality, location, etc.

At its simplest stage, this standard can be applied to analysis of the pure cost differences to the federal government involved in its choices of financial devices. This can be accomplished by applying the financial assistance parameters of each program in turn to a standard housing unit to determine the variations in net federal cost and net rent savings yielded by each of them. This assumes that the unit would have cost the same to construct and would have served the same tenants under each of the programs compared.

The next stage would involve the assumption that private construc-

tion cost of the comparable unit would have equaled that experienced by the assisted unit. By holding unit costs constant and applying the financial parameters of the going private mortgage market (regular FHA mortgage insurance terms and conditions may be useful for this purpose), we may approximate the net rental savings to tenants for each of the programs in turn.

Since, in either case, these rental savings accrue over an extended period of time, they must also be aggregated in some way to compare the relative "rental savings per dollar of federal cost" efficiencies of the various programs. Because these savings are effectively felt by the recipients as the (reduced) rent becomes due, the temptation is very strong to ignore discount factors and simply add the savings over the period in which they are expected to occur even though the costs would be aggregated on a discounted basis. The natural compulsion toward symmetry and reversibility in analysis argues otherwise. In addition, it may be useful in evaluating income-redistributional effects of programs to consider the stream of benefits as equivalent to a stream of cash payments. Thus, by discounting rental savings also at the federal borrowing rate, we have automatically calculated the present value of a stream of payments which can be compared with the present value of the pattern of federal costs incurred to yield that income benefit through an indirect (or multiple-goal) device.

Subsequent relaxation of the "equality-of-cost" assumption will be feasible through successive stages of architectural-engineering analyses of the products of each assistance program to approximate more accurately the private reproduction costs of assisted units. As sufficient operating and maintenance cost data become available, it will also be possible and desirable to put overall construction and operating costs on comparable bases to evaluate differences in total life ("systems"?) costs resulting from differences in construction standards as between programs.

In any event, the relative values of private "rental savings per dollar of federal cost" among the various federal assisted housing programs provide one benefit-cost ranking—one that measures the income redistribution benefit. We may term the highest benefit-to-cost ratio "the income redistribution optimum."

In addition to the income redistribution effects, there are other aspects of the effects of assisted housing programs requiring equal attention, in their turn, such as the unit output/cost optimum or "most bang for the buck" target.

Over and above direct financial cost/benefit comparisons, there are differential benefits associated with the separate housing assistance programs in terms of how each program contributes or fails to con-

tribute to the strategy of providing rebirth to dying cities or critical improvements to decaying neighborhoods. In short, if we were to give to the poor—through a negative income tax or some alternative device —the dollar equivalent of the income benefits they receive through these housing programs, we would still face the problem of saving our cities and neighborhoods; and this effort, too, would have its very substantial money cost.

Thus, should we identify two alternative housing programs which are comparable in their income redistribution benefits (or comparable in cost to the federal government), but one makes a notable contribution to urban redevelopment or neighborhood improvement and the other does not, then clearly the one making the dual contribution is to be preferred on efficiency grounds.

Accordingly, we would next want to identify the nonquantifiable environmental impacts of the alternative housing solutions and the long-term versus short-term implications. By and large, this would consist of identifying the favorable and unfavorable consequences worthy of note for each relevant housing alternative, including implications with regard to meeting other urban objectives such as the creation of sound neighborhoods and economically viable cities. We would also assess the relative significance of these factors, considering such things as the relationship to overall area needs in terms of roads, facilities, and commercial establishments; the impact on the neighborhood subculture; the hostility or acceptance of nearby residents; the ease of assembling suitable real estate; the probable timing; and the number of persons affected, etc. This could culminate in the creation of a table of descriptive summary data along the following lines: housing alternative; cost disadvantage (over least total cost solution); unfavorable environmental consequences; favorable environmental consequences; significance of consequences.

The final stages of this approach to an optimum solution would consist of a new ranking from "most preferred" to "least preferred" which considers both housing cost/benefit factors and environmental consequences. The theoretical optimum at this stage may be defined as one of the following: (1) the "least total cost" alternative whose net environmental disadvantages are not sufficient to offset its cost advantage over higher cost alternatives, or (2) the alternative whose net favorable environmental consequences provide the most significant offset (nonquantitative judgmental offset) to its cost disadvantage. We will call this approximation the "environmental optimum."

The suggested handling of nonquantitative data is not a mechanical or arithmetic one, and it does not yield a numeric solution. For that reason, it will be unsatisfactory to some. We are all familiar with the

horse and rabbit stew problem implicit in any decision involving both quantitative and nonquantitative data. However, we believe that this approach provides the opportunity to quantify everything that can be quantified in a practical situation and to compare systematically alternative packages of concomitant net benefits associated with each alternative as offsets to cost factors.

The President has placed strong emphasis on the careful definition of program objectives and on careful analysis of alternative means for achieving these objectives efficiently. At the federal level, the general strategy in housing and urban development programs will almost certainly be (1) to encourage efforts by local governments to identify and pursue the most efficient local solutions, (2) to develop national data and analyses which identify the most effective overall programs and administrative practices, and (3) to seek expansion of these efficient programs and general and broad application of creative new approaches. We invite your attention to these matters and seek constructive suggestions on the analytical methods we have outlined.

DISCUSSION

GEORGE W. HILTON: Unhappy is the lot of the discussant who thoroughly approves of the paper on which he must comment. What he had hoped in November would be a supercilious sneer proves to be merely another benign smile of the Christmas season. Rather than launch into fulsome praise of Mr. Enke's paper, I shall content myself with commending him briefly for the comprehensiveness of his treatment of the issues involved in federal participation in the supersonic transport program and for the nontechnical character of his presentation, appropriate to oral delivery.

Instead, I prefer to devote myself to the question whether participation in the supersonic transport program is an appropriate activity of the federal government. As Mr. Enke points out, it is generally agreed that the private sector of the economy will not bring forth the supersonic transport in absence of massive governmental support for the project. The most frequent justification of federal intervention one encounters in the press is merely that the investment is very large, but this is not an argument likely to satisfy most economists. The private sector has brought forth many innovations that entailed investments which were very large by the standards of the time, from the building of Pennsylvania Station to the construction of the 490-passenger subsonic aircraft currently on the drawing boards.

The reason the private sector does not bring forth a supersonic aircraft is quite clear. Speed, like any other service, is produced at increasing marginal cost. In this instance, marginal cost is discontinuous at the sonic barrier, beyond which control mechanisms, metallurgy, and other properties of the aircraft become markedly more expensive. Similarly, increases in speed yield diminishing marginal benefits. Replacing an 80-mile per hour railroad train with a 600-mile per hour aircraft cuts time between Chicago and Los Angeles from forty hours to four hours, but replacing the 600-mile per hour aircraft with a 2,000-mile per hour aircraft reduces elapsed time only another two and a half hours. This point is even more apparent in consideration of point-to-point travel. Development of a supersonic plane entails no improvement in transportation to and from airports. Indeed, the increased noise level of such aircraft, the necessity of aiming them where they are going, and the urgency of avoiding aerial holding patterns for them may well require the use of new airports in areas more remote from population. Abstracting from that problem and assuming that the supersonics use existing airports, successive increases in speed between given airports will result in successively smaller relative savings in time between specific urban points—a typical situation of diminishing returns from expenditures on a variable factor relative to a fixed one.

Accordingly, the explanation why the market mechanism fails to bring forth a supersonic transport is no different from the explanation why it fails to bridge Lake Michigan, fly us about in private helicopters, or do any of a wide variety of other things: the marginal cost exceeds the anticipated marginal benefit in each instance. The federal government typically refrains from inter-

vening in such circumstances, and so the question arises why it should do so in this case. It is not a situation, akin to the building of highways, in which the technology is such that private persons could not possibly collect the user charges. One might expect the reason for the intervention to be some external benefit, possibly to the military establishment, from the design of the supersonic transport. In this instance, the technology of supersonic flying has already been worked out for military aircraft, and the problem is to adapt the technology to commercial use. Any external benefits in knowledge will be serendipitous, of the sort which may arise in any technological change.

Indeed, instead of an external benefit, one finds a major item of social cost in the form of sonic booms. There is widespread expectation that the safety characteristics of the aircraft will be inferior to subsonic jets, and as Mr. Enke notes, aerial navigation control systems will be markedly more expensive. It is difficult to avoid the conclusion that one could make a better case for prohibiting or differentially taxing supersonic aircraft, even if economic, than for subsidizing their construction if they are uneconomic.

Presumably, the government's motivation in considering participation in development of supersonic aircraft is the same as it was in the building of the western railroads. That is to say, the government feels the project is economic, but its benefits are so far in the future that the market cannot take account of them. The recent researches of Robert W. Fogel and Albert Fishlow on the railroad industry in the nineteenth century should give pause to anyone using this argument. Fogel for the Union Pacific and Fishlow more generally have demonstrated that what were thought to be premature enterprises were economic from the outset. In the case of the supersonic transport, the uncertainties are undoubtedly less than in the instance of railroads into lightly populated western territory; thus, there is a stronger presumption that the market determination is correct.

J. M. Keynes once stated that the appropriate role for governmental economic activity was not assuming functions which the private sector provides, but rather doing what the private sector does not do at all. Economists would, in general, agree. In this instance, however, we have strong reason to believe that the private sector does not build a supersonic transport because it is not worth having.

ARNOLD MOORE: The five topics I will discuss are: (1) Growth of demand for long-haul travel. (2) The objective function to be maximized. (3) The need for recoupment via proportional pooling. (4) Organization of the U.S. government to handle this kind of program. (5) Possible alternatives to immediate SST development.

Dr. Enke notes that demand for air travel has grown rapidly in the past. This high growth rate has been caused in part by monetary price declines of between 1 and 2 percent per year.[1] Corrected for quality improvements and increases in the general price level, the price declines have been much greater; flight time, baggage handling, comfort, etc., have improved while price has

[1] Price is measured by "revenue per revenue passenger mile."

fallen. A rapidly expanding segment of the travel market is tourism. Tourists are especially price sensitive and seem to place relatively little value on their time. Continued falling prices may be a prerequisite for rapid growth. But the SST, which will require substantial market growth for program success, will find it difficult to compete in a market of low prices. Both the size of the supersonic surcharge and the absolute level of supersonic fares are important determinants of program success.

The objective function to be maximized in the SST program has never been made explicit. The SST program is supposed to be profitable, but it is also supposed to produce gold, employment, prestige and continued U.S. preeminence in subsidized civil aviation. Assuming that these are appropriate goals of government, the SST is probably not an efficient device for achieving them and may not achieve them at all. Large outlays beginning in 1967 are required for a balance-of-payments impact of doubtful size and direction beginning, at the earliest, in 1973. The aircraft and airline industries are booming without the SST program. It is doubtful that the SST will either create prestige or be the best way to "do something" for aviation, much less for transportation or for travelers.

Finally, government aid for the SST may be proposed simply because it is "too risky" for private development. But it has not been shown that the capital markets are unable to fund "worthy" projects of a strictly private nature simply because they are risky. The costs of adopting this as a reason for government involvement probably exceed any benefits.

Dr. Enke proposes that sharing (pooling) of all costs and revenues is necessary to provide appropriate incentives for the contractors to design the optimal commercial aircraft—an aircraft whose function is to make money from the airlines during operations rather than from the government during development.[2] I would urge that such a pool, if adopted, be either a single arrangement for both the airframe and engine manufacturers or a pool with the airframe manufacturer who in turn negotiates an agreement with the engine firm. Either arrangement can solve the bilateral monopoly problem Dr. Enke mentions. One of the reasons there is a problem in the relation between the airframe and engine manufacturers is that the government wants to be free to impose its own choice of engine on the selected airframe contractor. Such a pooling arrangement should be negotiated once and for all rather than following present practice of making new arrangements each time some technical milestone is passed. A financial plan should be based on economic decision points. The present procedure reduces the government's bargaining power and the contractors' incentives to commit resources. If a pooling arrangement is adopted, the contractors' cost shares must be set low enough that they will under all conditions be able to bear their share. This proposal is not a giveaway because the revenues are shared in the same proportion as the costs.

The proper definition of "cost" is crucial to the pool. The provisions for

---

[2] In discussing recovery of government funds, Dr. Enke notes that certain taxing schemes will disadvantage U.S. carriers relative to foreign carriers. These effects could be removed by writing the recovery provisions into the original sales contracts rather than making them "taxes" by the U.S. government.

sharing overheads, paying for specialized facilities, etc., must be such that the contractors' only source of revenue is the airlines. If the contractors can make money from the government during development, this may alter both the aircraft they develop and the development program, especially if the commercial prospects for the SST deteriorate. The amount the contractor puts at risk must exceed any hidden profit made on the contract or his contribution is meaningless. This creates pressure to raise the proportion of costs which the contractor must put at risk, countering the earlier stated requirement for a low cost share.

The pool is a device for encompassing the "externalities" and the "insurance" reasons for undertaking the SST program.

It would have been preferable if, from the start, the SST program had been viewed as a purchase by government of "externalities." Once a value had been placed on them—say, $300 million—government could have said, "Anyone who will agree to make an SST, introduce it by 1974, and sell, e.g., 200 to commercial airlines before 1985 can have $300 million." This offer could be left open with appropriate timely revisions of dates and amounts until accepted. This would have made the national interest cost explicit in advance and might have introduced an element of economic competition which is now lacking. It would remove some of the problems of cost definition and of government's commercial judgment which are likely to plague the program.

A third major topic is how U.S. government should organize itself to procure good advice on such projects and carry them out appropriately. The major U.S. involvement with the SST began with a speech by President Kennedy. While his remarks contained appropriate caveats about stopping if the program develops unfavorably, they have been interpreted as a strong go-ahead. Observation of the SST program suggests several generalizations.

1. Programs of this sort are usually generated by fascination with some technical novelty. Detailed economic analysis comes late, if at all. Further, it is often done by the interested agency on the theory that close knowledge of the "industry" and the "technology" are more important than detachment. Thus a prime need is for introducing dispassionate rationality earlier. This could be introduced by the Bureau of the Budget's Program Evaluation Office or the CEA Staff. Early involvement is very important. Merely choosing among the designs that engineers dream up is not sufficient. Economy must be the design criterion. The SST program would have benefited greatly from such early analysis as would the programs for nuclear merchant ships and possibly nuclear electric power and communications satellites.

2. The most important single thing that could be done to facilitate analysis of such proposals would be to avoid any public comment as long as possible. Favorable statements by the President make it difficult for government agencies to entertain any but the most enthusiastic notions. Indeed, once such public positions have been taken, it becomes increasingly difficult for the President to get good advice.

In addition to these general comments some observations apply specifically to the SST program. In the SST program, conflicts among the government's goals and duties exist.

1. The federal government must certificate both the U.S. SST and the Anglo-French Concorde. The conflicts here are numerous and obvious.

2. The federal government may also become responsible for establishing standards for public acceptability of levels of sonic boom and airport noise. Here, the temptation to make the standards fit the plane will be great and has already shown itself.

3. The government must also set airline fares. There is an obvious conflict between the CAB's laudable desire to reduce fares and the harsh impact of low subsonic fares on the SST program. Perhaps within the government these conflicts will be resolved with due regard for the public interest, though precedents do not suggest it.

Finally, if the SST does not look economic now, are there any real alternatives to pressing ahead given the apparent intent of the Anglo-French Consortium to continue with the Concorde and the threat Concorde is alleged to pose? What can be done to moderate our pace by moderating the prospects of the Concorde for early success? First, the U.S. government can reduce its funding of programs which help both the SST and the Concorde. These are the development of air traffic control and multi-billion dollar international high altitude weather reporting systems required for safe and economic commercial supersonic flight.

In addition, the government can push very hard for reduced subsonic fares —desirable on other grounds. It could also push for expanded research in public reaction to sonic boom. Together these actions would take much of the sting out of the Concorde challenge. They would very probably make possible a more deliberate and more nearly commercial program for developing and introducing an SST about five or ten years later than is now contemplated.

MARTIN ANDERSON: The basic question raised by the three preceding papers is, how shall we evaluate the actual or proposed operation of certain government programs? Implicit in the question is the possibility that a program may be found undesirable and stopped. Thus we are faced essentially with the question whether or not certain programs are justifiable. The basic approach suggested in these papers is to measure benefits and costs, to compare them, and then to decide. It is implied that if benefits do not exceed costs, the programs will not be undertaken or continued. While this is debatable, let us assume it is true.

The cost-benefit approach seems plausible enough and has the added appeal of apparent objectivity. But while I agree with the general cost-benefit approach to economic problems, I am afraid that the basic methodology proposed here is incomplete and, in significant aspects, conceptually incorrect.

My criticism lies mainly with the basic concepts employed in developing the methodology. The proponents of any methodological approach must start with certain basic concepts, and the validity and usefulness of the result depends directly on them. The concepts themselves must be valid and precisely defined.

The key concepts in the methodological approaches suggested here are benefits and costs. The programs mentioned involve many people and it is im-

plied that we can somehow add up the benefits enjoyed and the costs borne on a collective basis. But can we meaningfully talk about collective benefits and collective costs, or are these just vague, nondefinable terms used to bridge gaps in the logical development of a methodology?

Let us first examine the nature of a benefit and a cost. Both concepts are derived from the concept of value. A value is something which a person acts to gain or to hold. A benefit is the receipt or retention of a value. A cost is the giving up of a value.

Epistemologically speaking, only an individual can value something. It is only conceptually correct to speak of values from the viewpoint of a single person. There are no "group" values, no "collective" values, no "society" values. The concepts of group or society are abstractions which refer to collections of individual entities. Only the individual entities themselves are capable of valuing and of receiving benefits and incurring costs.

Of course, we often refer to groups of individuals as if the group actually did value something. For example, we might say that a certain school values the teaching abilities of its faculty. But we continue to recognize that this is just a shorthand for saying that certain individuals within the school value teaching ability. And, of course, it is possible for people to share the same values.

A proposed methodology for evaluating the costs and benefits of a program which directly affects the lives of many individuals is a situation where collective concepts are inappropriate. At best it represents nonrigorous, imprecise thinking which obscures what is actually happening. A correct methodological approach to the evaluation of government programs would have to ask such questions as: (1) Who benefits and how much? (2) Who pays or gets hurt and how much? (3) If some people benefit and some incur costs, is it proper to deliberately sacrifice the interests of some people so that others may benefit?

These questions have been virtually ignored in the proposals put forth. The question whether or not some people should be sacrificed in the interests of others is, of course, a moral question, and it is certainly relevant to any evaluation of a government program. But before we can even consider it, we must clearly identify what is happening. Vague comparisons of "total benefits and costs" simply obscure the incidence of the benefits and costs.

An example of a government program where such an analysis is particularly relevant is the federal urban renewal program. The federal urban renewal program has the distinction of actually worsening the situation it was supposed to help. Started in 1949, it is probably the major weapon in the government's attempt to achieve the expressed goal "of a decent home and a suitable living environment for every American family"—a goal of which Mr. Ross extolled "the directness, simplicity and grandeur."

And what has happened? Billions of dollars of the taxpayers money have been spent on one of the most massive and cruel fiascoes of our time. This program has forcibly evicted over one million people from their homes. Virtually all of these people are poor; two-thirds are Negroes and Puerto Ricans. The program has succeeded in destroying tens of thousands of low-rent homes and in building primarily high-rent, luxury units. It has worsened housing

conditions for those whose housing conditions are worse; it has improved them for those whose housing conditions are best. The beneficiaries are those who can afford the average rents of $200 or more a month. Who was hurt? Who paid? The poor, the minority groups, and the taxpayers.

These are certainly relevant issues and a methodological approach which ignores them is incomplete. But then these are issues which many people find unpleasant to discuss, for they make it clear who is sacrificed to whom and for what reasons.

In fact, all the government programs considered here involve the deliberate sacrifice of some people to others. For example, our government is now considering a supersonic transport which will cost the taxpayers billions of dollars. Who will benefit? The very, very few who can afford the high fares, which would be astronomical if it were not for the fact that the nonflyers will pay a substantial part of the cost. As Mr. Enke points out, only 15 percent of United States residents have ever flown, and perhaps less than 5 percent will ever fly on the supersonic transport. The other 95 percent will have to be content with the "prestige."

In summary, there are no collective goals, no collective benefits, no collective costs. There are only individual goals, benefits, and costs. It is time to abandon fuzzy concepts which are only euphemistic devices for masking the essence of what is being done. We could take one small step away from a "Great" society toward a "Just" society by clearly identifying and recognizing the costs and benefits which fall on individuals.

HENRY S. ROWEN: Worth Bateman has presented an analysis, both methodological and substantive, of one of several federal poverty programs, HEW's Work-Experience Program. His paper makes it explicit that an adequate analysis of this program is not possible at this time. This is so partly because the program is too new for enough data to have been generated and partly because in this program as in so many government programs, there has only recently been much interest in systematically gathering data.

Most of my comments, therefore, have to do with methodology. And perhaps the best place to begin is with the subject of the antipoverty effort.

This effort is not based on a single coherent view of the phenomenon of poverty. Far from it. It is, in fact, one of the real strengths of the present poverty program that it is so diverse and experimental in character. Out of the welter of approaches being taken, I would like to identify: (1) the aggregate demand approach; (2) the subsidized jobs approach; (3) the public assistance approach; (4) the money transfer approach; and (5) the education-training approach. Now clearly it is a combination of these approaches that is of greatest interest to government. For example, by now everyone agrees that, whatever else is done, there should be the right level of demand; but there is real disagreement on what is the right level.

There is fairly widespread agreement that the structural problem is serious at about 4 percent level of unemployment, especially in rural areas and Negro ghettos. There is also agreement that the aged poor need money, not jobs or education. But beyond this point, the divergence becomes fairly great.

Part of the difference lies in the population group one has in mind: youths

versus adults; rural versus urban dwellers; black versus white; single versus head of family; people with a basic education versus illiterates; mentally retarded versus normal intelligence. One of our difficulties is that we either do not have or have not exploited enough structured data for such groups.

Now the main premise on which the Work-Experience Program is based is that there exists a sizable number of people unable because of skill limitations and market imperfections to compete in the labor market or not sufficiently motivated to try, who, if given some guidance or motivation or minimal skills could become productive—although not necessarily at the federal minimum wage level. Of course, in some cases, one of the reasons for lack of motivation may be the availability of public assistance support and to the extent this is so, it provides an instance, unhappily not unique, in which one government program is designed, at least in part, to offset the effects of another government program.

Now how are we to judge the efficiency of this program. Bateman emphasizes economic benefits measured in terms of earnings. Although these are undoubtedly important, I wonder if some other benefits are not more important —at least for some groups. For example, as Bateman mentions in passing, the effect on family stability, on children, on reduced delinquency and crime, and on working versus being idle may be quite important. At least for youths and young adults. If so, it may be worth supporting the program, even if program costs exceed direct earnings increases. The main point, however, is that we do not know at this time.

However, the prospects for finding out enough to make reasonably based decisions are easier than in many other areas of governmental concern. One can experiment on a fairly small scale and probably obtain valid results within a few years. Contrast this with such a large, "lumpy," long gestation-time investment as the supersonic transport or many aspects of defense where, one fervently hopes, some kinds of experimental evidence will not be forthcoming.

Finally, I would like to comment on the problem of handling the large numbers of variables that seem relevant to poverty: education, work skills, housing, transportation, health, motivation, family size, and so forth. One would like to be able to isolate a single key factor, attack it, and solve most of the problems. It seems implausible that such a factor exists. It seems more probable that much of the environment must be changed. This is the Job Corps approach, to take youths out of their old environment and put them, for a period, in a new one. But the old environment is complex and we cannot change everything in it, nor can we, nor should we, send everyone to a Job Corps camp. Nor do we seem to have very effective bureaucratic tools for identifying the key variables and bringing about the right kind of change in them. For example, there is reason to doubt the efficacy of local welfare institutions in doing this. Efforts are being made, as Bateman points out, to develop multi-component projects. It is important that we gather data in a way that will make it possible to do multivariable analysis. I do not mean to suggest that it will be easy, but government bureaucrats are used to hard problems.

ECONOMICS OF HEALTH

THE ALLOCATION OF BIOMEDICAL RESEARCH*

By Simon Rottenberg
Duke University

Medical research can be treated as an industry. It consumes resources that have alternative uses and that are scarce. It can, therefore, be of a suboptimal scale; it can be too large or too small.

The optimal scale is that which satisfies the textbook condition that marginal cost and marginal revenue be equal. The calculation of revenue requires the operational definition of the product of the industry. Research is an activity that produces incremental flows into the stock of knowledge. Medical research produces flows into that part of the whole stock of knowledge, which, when combined with other inputs in other, complementary industries (say medical care or public sanitation) produces "better health." The immediate product of the medical research industry (i.e., more knowledge of the relevant class) can be treated as a cheapened input of the complementary industries or a more efficient input acquired at given cost. Thus, with a given quantum of resources employed by those industries, an increment of output of the medical research industry can operationally be thought to produce, say, one less illness (adjusted for the mean duration of the illness, for the probability of disablement of those who suffer the illness, and for the probability that the illness, once suffered, will be terminal); or reduction of the mean duration of an illness by one day; or reduction, by some magnitude, of the probability of disablement of those who come down with the illness; or reduction, by some magnitude, of the probability that, once suffered, the illness will be terminal; or the prolongation of life for x months (adjusted for the number of dependents, income); or the increase, by some magnitude, of the productivity of functioning persons (as by diminishing the frequency or severity of asthmatic attacks); or the reduction of the pain associated with an illness by one disutil.

Optimality of scale means, illustratively, that the industry is that big such that the cost of the extra resources used to prevent the extinction of the last life shall be equal to the value of that life. Or that it shall be that big such that the marginal cost of repairing life (i.e., restoring from illness, diminishing the duration of illness) shall be equal

to the value of the fraction of that life with respect to which restoration occurred.

Decisions with respect to the size of the medical research industry involve either explicit or implicit choices among objects desired by society. These decisions are responsive to the question: what is it worth society's while, in terms of other desired objects foregone, to prevent, say, the extinction of a life. Decisions with respect to the allocation of medical research activities among diseases also involve choice. Diseases have different coefficients of incidence for different classes of the population. Men are two to three times as vulnerable to peptic ulcers as women. The incidence of cancer of the uterus is positive for women and zero for men; cancer of the stomach falls and cancer of the intestines rises with age. Assuming a first-approximation premise that the quantity of information produced by research is a direct function of the quantity of resources employed by the research industry and given that research resources are scarce, it follows that the allocation of medical research among diseases involves choice among lives. Anything done to increase the probability of survival of a set of which John Doe is a member will cause the survival probability of the set of which Richard Roe is a member to be less than it otherwise might be.

Lives that are saved from extinction are conventionally valued by discounting, at some rate, the expected earnings of the saved life over the expected duration of working life, adjusted for the probability of survival to the age of retirement and for the probability that "he" will be employed. There is, of course, a considerable literature on whether the consumption of the "deceased" should be netted out. To this should perhaps be added the value of the preference for the "state of living" over the "state of death" and the negative value of the grief suffered by survivors if death occurs.

While all of these components of valuation are relevant to the decision with respect to the optimal scale of the medical research industry, an assumption that all value equally the preference for living and that all survivors grieve equally makes the last two components irrelevant to the allocation of research among diseases. This assumption does not, however, seem to be consistent with what we observe in the world. Some are willing to put their lives at risk at a cheaper price than others; some look both ways before crossing streets and others do not; some eat pigs' feet sitting on railroad tracks and others sit in safer places. Similarly, even if grieving were equally intense for all, the quantity of grief produced by a death would be proportional to the number of survivors which varies among the dead; and grieving is not equally intense if, for example, the length of run between loss of a husband and remarriage can be taken as a measure of grief's intensity.

Alternative measures of the value of life can perhaps be constructed by manipulating data on the differences in payments made for services in occupations with different risks or data on the prices of safety equipment or on the amount of life insurance carried.

If the value of life is calculated according to the conventional method, the values of some lives might turn out to be negative and all positive values are finite.

In any case, different subsets of the population can be expected to turn out to have different values. The cost that it is worthwhile to incur to prolong life is a partial function of prior investment in the formation of human capital. Other things being equal, since society loses more from the extinction of more highly valued lives, it would pay it to devote more resources to the preservation of those lives than to less highly valued lives. *Ceteris paribus*, therefore, it pays to use more resources in medical research directed towards diseases that strike heavily upon more highly valued population subsets and less resources oriented to diseases that incide upon less highly valued subsets.

Mankind is a durable good subject to physical depreciation the rate of which can be kept in check by maintenance and repair expenditures. On economic grounds, approximately the same rules that govern when to maintain and when to scrap machine capital and that determine the optimal rate of maintenance expenditures for machine capital also apply to human capital. These are different rules from one which would allocate medical research among diseases according to their rank order of incidence of morbidity and mortality. They are also different from the rule that says: give particular attention to diseases found in the age class that comprehends a large fraction of members of Congress or found in Presidents or Presidents' families. Such a rule tells us that society will value more highly a child with polio than a child with leukemia because a President is afflicted by polio and indeed it suggests to voters that, in choosing among candidates, one of the criteria of choice should be the coincidence of the voter's diseases and the candidate's.

Although laymen will usually believe that the value of human life is infinite, the aggregate of them behaves as though it is a finite number. There is a paradox, however, in observed behavior. If an octogenarian alcoholic threatens to jump from a bridge, enormous resources of passers-by, priests, police, and firemen are mobilized to prevent his extinction. But if the life which is threatened is not a certain, known one but rather an unknown, probabilistic one, we are much more sanguine. Thus the probability of fatal accidents at curved roads is higher than at straight roads. Lives could be saved by straightening the road but to do so would deprive society of the utility derived from seeing films in a

movie theater. Since we do build theaters, we behave as though the utility produced by the seeing of the film is worth more than the saving of some number of lives by the straightening of the road.

The gain from medical research includes not only the capital value of income streams preserved, but also the cost of therapy saved as, for example, when diseases are prevented or illnesses shortened.

Gains can be seen in another context that involves allocative choice of an order not yet mentioned. The preservation of life and the prevention of illness can occur through other processes than medical research. It can be done also by straightening roads, patrolling them more intensely, designing more secure elevators, building fireproof buildings, preserving food, adding nutrients to foods that they do not now possess, finding new and palatable sources of protein, bathing or brushing teeth more (or less?) often, moving people away from flood plains, producing more or better physicians, etc. If it is known that the scale of the lifesaving and life-improving industry should be x, optimally, there still remains the problem of minimizing cost for the given output by allocating resources among the enumerated (and other) subindustries. Medical research is only one way to skin a cat. In calculating the gains from the prevention of a disease there are two other problems: preventive measures may not be single-disease specific and this creates an accounting problem; and the prevention of one disease increases the probability of contracting another. It is not clear whether netting should be done in calculating gains.

We turn now to a discussion of some of the properties of research on the cost side that affect allocative decisions.

If the relative magnitudes of gains that would be produced by the acquisition of different increments of knowledge were calculated, it would still be necessary to estimate the costs of acquiring different increments in order to know which of them are more, and which less, worth seeking.

The literature of discovery in medicine is larded with cases that exhibit the fortunate consequences of accident. *Rauwolfia* alkaloids were introduced to treat high blood pressure; "quite accidentally" it was found to calm excitement and this led to the use of its most active constituent—reserpine—as a tranquilizer for mental patients. Ringer's laboratory assistant was bored with using distilled water and used tap water instead in making physiological saline. From this it was discovered that it is valuable to add calcium and potassium to physiological media. Gram found how to stain bacteria by taking the wrong bottle from a shelf. The use of iodine in the treatment of Grave's disease was discovered when Trousseau "absentmindedly" prescribed tincture of iodine for a patient when he had intended tincture of digitalis.

From incidents such as these there has been constructed a notion, which is widely held in the biomedical community, that research is a serendipitous undertaking. If this were true and even if outcomes—say, a reduction by 40 percent of the current incidences of a number of different diseases—can be ranked in the order of their capitalized present values, there would be no guides for the allocation of research resources.

But, of course, if serendipity were really believed to be as important as it is often said to be, we should observe research fund granting institutions using a random number generator to choose among applicants, and researchers would also use random processes in constructing research designs. If slipshod methods were so fruitful, we would see research designs deliberately prepared to generate mistakes. If assistants are not blindfolded as they move abut their laboratories and if researchers who desire to discover A are not observed to search for B and C, it is perhaps because the biomedical research community does not really believe accident is all that powerful.

Nor does the researcher stand in pristine innocence, waiting to see what his method will reveal. At the least, implicit hypotheses lurk in the background and tests are contrived that seem appropriate to them. Hypotheses thought a priori to be inconsistent with the world are rejected out of hand; they are not tested. But if an hypothesis is tested, it is because the subjective probability assigned by the researcher to its consistency with experience is large. That is to say, he expects the outcome to occur.

This is not to say that biomedical research is not uncertain. It is, and in two senses. Hypotheses, upon test, might be rejected; but this is like firms that might be bankrupted and research is not especially vulnerable in this respect.

More importantly, uncertainty derives from the characteristics of the subject matter of the research. Some biomedical research is oriented to particular diseases (like clinical studies of ulcerative colitis of the biological control of schistosomiasis). Much, perhaps most, of it, however, is in the basic sciences such as biochemistry, microbiology, biophysics, physiology, and anatomy (for example, studies of the general metabolic effects of hormones or biochemical and genetic studies of mammalian cells). The outcome of such research might not be very uncertain with respect to its immediate object (the production of an increment of knowledge) in the sense that a set of different possible outcomes, each with a subjective probability estimate attached to it might not be characterized by large variance. The revenue-producing applied outcomes which are generated by the use, by other industries, of the knowledge produced by research, are not, however, known *ex*

ante. The distribution of this set of applied outcomes does, apparently, have the property of large variance.

To choose in conditions of uncertainty is not to be reduced to perfect ignorance and random or arbitrary rules of choice. Rational choice rules still have relevance. It is estimated that $1.85 billions were spent for medical research in the United States in 1965. Of this, $1.24 billions were provided by government and $.45 billions by the pharmaceutical industry. Perhaps 5 percent of all health and medical expenditures is spent for research. Elaborate processes of review shaped the structure of research that was undertaken. The review panels and other review arrangements taken together might, in principle, have been badly instructed; the priorities given them might have been badly ranked. Whether or not this is so, however, given their collective terms of reference, their examination of the promise of proposed researchers and proposed research designs, in terms of the objects sought to be achieved, constitute the implicit execution of probability estimates on how far the present state of knowledge is from the threshold of discovery. This is confirmed by the fact that funds allocated to different projects are of different magnitudes. That some are given $20,000 one-year grants and others $200,000 five-year grants implies that something is known about the comparative distances, in terms of the quantities of resources that must be employed, from the achievement of intended discoveries.

Biomedical knowledge progresses in stages. One distinguished biochemist has suggested that it begins with inquiry into the nature of living cells and is followed, sequentially, by investigation of the physiology of the mammalian organism, laboratory investigation of experimental disease states, investigation of disease as it occurs in man, studies of the mechanism of drug action, experimental trial of new drugs, development of instruments for use in diagnosis or clinical care, clinical trials of therapeutic materials and, finally, clinical care by accepted procedures. This suggests that it is possible to assign probability numbers to research outcomes: research on stage six is more probably successful if prior research has carried to five than if it has carried only to two.

Where uncertainty means that discoveries intended to apply to one disease have spillover effects for others (as benzedrine, developed as a nasal decongestant, was found to be useful for treating mild depression), this can be taken care of by the upward adjusting of expected gains by some factor in calculating the quantity of resources it is worthwhile to use for the intended object.

The allocation of research activities among diseases is not by a long shot the only aspect of this industry and related industries with respect

to which there are decision-making problems. It is a field permeated by opportunity costs, and, therefore, it contains abundant possibilities for the application of rules of rational choice.

Better health will be achieved, not only by increasing the stock of medical knowledge, but also by funneling the currently available stock into the hands of more practitioners and by motivating people with incipient or first-stage ailments to submit to medical treatment, given the knowledge that practitioners currently command. To which of these competing objects should an increment of resources be best put, given relative costs and gains? Physicians are said to be casual in the administration of anticoagulants for forms of heart disease. Would a unit of resources be better used to reeducate them than to do research? Or should it be used to inform middle-aged males that they are prone to coronary attacks and to spread the word about "the seven danger signals of cancer"? Or should it be put to the improvement of sanitary works, or to rehabilitation, so that the infirm can function as closely as possible to the limit of constraints imposed by their infirmities? If some environmental event is associated with a disease, is it better to find a preventive such that that event does not produce the disease or to change the environment?

Within medical research, even when allocation among diseases has been worked out, there are additional allocative problems. What kind of knowledge should be searched for? Prophylaxis? Improved diagnosis? Or improved therapy?

Where diagnosis now requires multiple symptoms, simpler diagnostic signs can be sought, as, for example, a blood pressure reading is a good predictor of later cardiovascular disease. Streptococcal infection generates clinically-observable symptoms that cause diagnosis to be suspected, but it is necessary to perform laboratory tests for bacterial confirmation. Should research be done to cheapen the cost of diagnosis by seeking to discover additional clinical symptoms that would make laboratory testing superfluous? Should subclinical and silent symptoms be caused to become overt? Should research seek to raise reliability estimates in diagnostic screening? Reliability estimates that are lower than one cause some diseases to be undetected and cause some who are free of the disease to submit to therapy. Only about 3 percent of persons suffering from certain streptococcal infections come down with rheumatic fever. The options are to administer relevant prophylaxis to all (or none), or to attempt to discover the distinguishing characteristics of those who are and are not susceptible to rheumatic fever. There are trade-off opportunities between the cost of therapy and the cost of discovery. Syndromes may define a class of etiologic agents but not a specific disease. Therapies will be different for specific

diseases, if they are identified in a particular case, than for a class of diseases of which the specific disease is only one member. Even if no therapy is known for a given disease (that is to say, administered therapy produces a cure in zero percent of the cases), better diagnosis saves the cost of misplaced therapy.

If it is improved therapy that biomedical research seeks, shall it be for cures that diminish the intensity of an illness (that reduces the mean duration of illness) or that reduce the ratio of terminal to nonterminal cases for those contracting the disease or that reduce the cost of complements in the application of the cure? Therapeutic procedures that are equally effective and that do not require the services of medical practitioners are probably preferred to others that do; those that permit practitioners to treat patients quickly or a number of them at one time are preferred to slower procedures and those requiring unitary physician/patient ratios; those that permit outpatient treatment are preferred to others requiring hospitalization; those using sulfadiazine are preferred to those using oral penicillin. Or should the search be for therapies that diminish adverse side effects. Patients with acute heart attacks have been administered fibrinolysin to dissolve clots. But this agent has the disadvantage of causing antibodies to it to be produced and it also produces pyrogenic reactions. Congenital heart disease is sometimes produced by certain medication of the mother during pregnancy. Given the existence of opportunity costs, would it pay off to seek other agents that do not have side effects? Smallpox, malaria, typhus, typhoid, and other fever diseases inhibit paresis as a late manifestation of syphilis. To eliminate the fevers is to induce paresis. Where should research confronting this problem stand in relation to others in the competition for research resources? Therapeutic outcomes can be various and each has an associated probability number. Should research seek to increase the probability numbers associated with preferred outcomes?

Or, again, should research be designed to effect discoveries that will reduce the incidence of recurrence of a disease, as those which brought down the recurrence rate from ulcers from 80 percent at the turn of the twentieth century to about $\frac{1}{2}$ of 1 percent now when surgery is properly done.

It is assumed, of course, that research resources are not perfectly specialized; if they were, the only choice would be between idleness and their employment in the single use of which each is capable. It is assumed also that technical coefficients are not fixed and that there are alternative research designs for achieving given objects. The construction of least-cost factor combinations in the doing of research is still another allocative problem. "Little science" research competes for se-

lection with projects requiring large primate colonies, computers, and large clinical samples, and mycobacteriologists compete with clinicians.

Two other considerations affect the execution of choice among alternatives in biomedical research.

One is that while knowledge, once produced, has a "physical" life that runs to infinity, it is subject to technological obsolescence. The relevant rate of obsolescence is affected by the distribution of investment among research projects according as these investments seek to discover complementary knowledge or substitutes for given knowledge.

The other is that discounting affects rational research choices in a number of contexts. The present value of the prolongation of life into a next year is greater, other things being equal, than the present value of the prolongation of life into the twentieth year hence. *Ceteris paribus,* therefore, research that will prolong the lives of twenty men for one year is to be preferred to research that will prolong the life of one man for twenty years.

Every increment to the stock of knowledge will have a time-payoff pattern. These patterns can have any shape. They can rise over time or fall. They can rise from a zero payoff quickly or only after a longer period has run. If the patterns of payoffs can be foreseen, it should enter the calculus of choice among projects, with delayed gains discounted. On the cost side discounting has an opposite effect. Research projects can be perceived to run to a time-point of discovery. Other things being equal, projects involving deferred costs are to be preferred to projects involving costs in the present; the longer the deferment, the better.

Neither multiplicity of options and variables nor variance of outcomes transforms biomedical research to a condition which makes rational decision making irrelevant. The whole rationalizing apparatus for determining any other kind of investment can, in principle, be applied to investment in the discovery of incremental bits of knowledge.

The fact that many decisions in this field require the estimation of probability sets does mean that mistakes will be made. Perhaps the coefficient of error can be held to a minimum by subjecting government research fund grantors to a discipline simulating that imposed by the market. There private investors are differentiated according to their capacities to estimate objective probabilities and payoffs correctly. Those that make good estimates survive; others go to the wall. Could not *ex post* evaluation be made of the judgment of panels and panel members so that it could be seen which of them systematically recommend approval of research projects that do not pay off, so that they shall not survive?

That rational rule making, as economists conventionally define it, can govern does not mean that it has. Often allocative decisions in biomedical research seem to have been responsive to one or another of two other rules: either "let the researcher be funded to do what interests him," or "let knowledge be found that we do not now possess."

Such rules are congenial to the scientific community and they are sometimes defended on social optima grounds. It is said that more that is useful will be (has been?) found following these strategies of choice than would be (would have been?) found, if, say, public monies were made available to research according to the calculus of the relationship between the cost of discovery of a bit of knowledge and the value of the gain that its discovery will generate. This conclusion is based on what is asserted to be true of empirical experience, on the claim that researchers know better than grant strategy calculators how great are the distances to the thresholds of discovery of different increments of knowledge, and on the belief that productivity in research is adversely affected if researchers are enticed, by the availability of funds, to fields that are ranked low in the hierarchy of interests ordered by the idiosyncratic properties of their curiosities.

If the researchers laid claim to public funds because they desired to produce aesthetic utility for themselves, their position would be without merit. Since their rules are offered as strategies designed to promote social objects, they are deserving of attention, but they are likely to be less effective in the service of that purpose than others that take more explicit account of opportunity costs.

ECONOMICS OF HOSPITAL SYSTEMS: PEAK LOADS AND REGIONAL COORDINATION*

By MILLARD F. LONG, *University of Chicago*
and PAUL J. FELDSTEIN, *University of Michigan*

The demand for hospital services fluctuates from day to day. To meet these variations hospitals must build larger facilities and hire more staff than would be necessary if demand were constant over time. We are presently investigating techniques for calculating optimal facility levels for short-term, general hospital systems, given the "peak load" nature of demand.[1] This work is incomplete and in any case would be too long to report on here. In this paper, therefore, we shall concentrate on one service: obstetrics. The two reasons for choosing obstetrics are, first, that the statistical characteristics of obstetrical service and demand make it possible to cast the analysis in a rather simple mathematical form, and, second, the problem of low occupancy is greatest in obstetrics.[2]

One alternative to the present organization of hospitals would be a system of profit maximizing, proprietary units.[3] In such a system hospitals would charge higher prices to ration places when crowded; market signals would then indicate the desirability of expanding or contracting facilities.[4] Instead of maximizing profits, hospitals seek to optimize some complex, differing, and, for most institutions, ill-defined goal subject to certain financial constraints. Commonly, hospitals attempt to "break even" on operating account.[5] For capital sums, they depend upon government contributions and charitable solicitations. Hospitals do compete with one another. With a limited loss constraint rather than the goal of profit maximization, however, there is no assurance that separate institutions acting independently will produce a so-

    [1] For another discussion of the same problem, see M. Long [16].
    [2] In the Chicago region, for example, the percentage of available bed days actually utilized in 1965 was 61 percent for obstetrics, 71 percent for pediatrics, and 84 percent for medical-surgical beds.
    [3] Externalities, community selected special services, and charity patients could be handled through subsidies.
    [4] Admittedly there are difficulties in using prices to ration crowded facilities. For a discussion of how pricing could be used to deal with some of the problems discussed in this paper, see B. Weisbrod [22].
    [5] In some instances the hospital will have a sponsoring agency which will specify the amount the agency is willing to contribute annually towards an operating deficit.

cial optimum. Realizing this, the major contributors of capital have urged all states and many of the larger cities to engage in regional planning to minimize uneconomic duplication as some coordination of activities seems socially desirable.[6]

Rules of thumb are usually used to translate demand estimates into facility requirements.[7] In this paper we attempt to apply more formal tools to the problem. To judge whether the number of units and the number of beds are optimal, we suggest the following criterion: for a given number of obstetrical cases[8] costs should be minimized, including the costs of hospital care, travel[9] and inconvenience. The last category includes an assessment for reduced quality of care and inconvenience to community members when the preferred facilities are not readily available.[10] Unfortunately, the difficulties inherent in measuring the components of this function make exact marginal calculations impossible. Our approach indicates the range within which the correct answer probably lies.

The long-run relationship between average costs and size of an obstetric facility is depicted in Figure 1. The hospital cost curve reflects economies of scale, both because over a range there are indivisibilities of staff and equipment,[11] and because larger units operate at higher occupancy levels. Higher occupancy rates make possible more intensive use of facilities and staff and by spreading the overhead reduce the cost per case.

Larger units have higher occupancy rates because the populations they serve are larger, and the larger the population the lower the coefficient of variation of demand.[12] With a lower coefficient of variation, fewer beds (relative to mean size) must be provided to meet peak

[6] Each state has a Hill-Burton agency in charge of regional planning; without the approval of this agency no federal funds can be obtained for hospital construction. Approximately forty cities, including almost all the largest, have hospital planning councils of their own. See J. Cavanaugh [6] for a discussion of the growth of areawide planning.

[7] See for example, Joint Committee [25] and A. Airth and D. Newell [1].

[8] Today, better than 96 percent of mothers have their babies in hospitals. Therefore, we have treated demand as given without concern for feedbacks between costs and demand. Length of stay, however, may be influenced by price variables and a more complete planning model would take this into consideration.

[9] For a discussion of the effects of travel on hospital usage, see Schneider [20] and P. de Vise [11].

[10] Regardless of whether compensation is paid to those inconvenienced, penalty costs should be included in the cost function. Were the solution profit market determined, those who wished to avoid inconvenience could do so by paying a price. With planning the costs may be allocated differently, but the level of facilities should be the same. See R. Coase [7].

[11] The indivisibilities are both technical and legal. Most states have laws specifying the minimum staff that must serve on a maternity unit. For a discussion of indivisibilities on obstetrical wards, see Hospital Review and Planning Council [26, pp. 17-23].

[12] It is a common error in the literature to associate the variation in demand with the hospital size rather than the population size. This error leads to the conclusion that small hospitals, because of the large variance in demand, will have low average occupancy levels. Of course, this is true of small hospitals which serve small populations, but it need not be true of small hospitals in large urban areas where the population is served by more than one hospital.
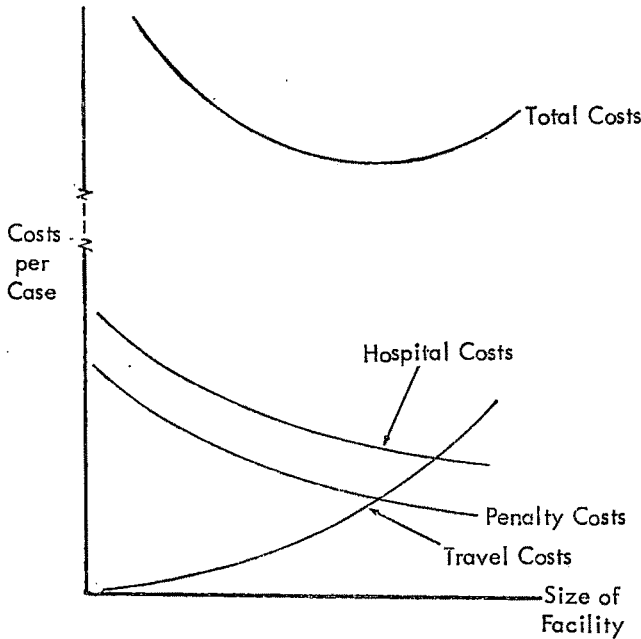
FIGURE 1

loads which implies higher average occupancy in larger units. To illustrate, think of two small hospitals handling the same number of patients as one large one. With two hospitals, one may be full while the other still has unoccupied facilities.[13] Lack of correlation among peaks reduces the relative variation when the two units are combined. A key conclusion of this paper is that higher occupancy rates can be achieved without physically combining units.

In Figure 1 travel costs are shown to rise as units get larger because larger units imply fewer hospitals and hence patients must travel further for service.[14] The third category of costs, penalty costs, are incurred by those inconvenienced when the absence of facilities necessitates a departure from regular procedures. Hospitals do not use prices

[13] Occupancy levels in different hospitals tend to move together, but the movements are not perfectly correlated. On rare occasions, as happened in New York City nine months after a power blackout, all maternity units may peak simultaneously. Except in the case of perfect correlation less beds, relative to the mean, will be needed in larger units to reduce the probability, say to one in 10,000, of a patient arriving to find all facilities full. For example, suppose x_1 and x_2 are jointly normal variables with common mean, μ, variance σ^2 and correlation coefficient ρ. Then Var $(x_1 + x_2) = 2\sigma^2(1 + \rho^2)$. Except in the extreme case ρ^2 will be less than 1.

[14] With units of given size, more travel will be necessary in less densely populated areas. Optimal unit size, therefore, will probably be larger in metropolitan areas. To simplify the drawing of Figure 1, I have assumed a uniform population density, so that travel costs can be represented by a single function. This means one optimum size for all hospitals in the region. In considering the existing distribution of hospitals in Chicago, this assumption was dropped.

to ration peak capacity. Instead they stretch capacity by crowding existing facilities; or they release beds by dismissing some patients early; or they delay elective admissions until facilities are available; or they substitute other facilities for those in use, e.g., they put a child in an adult ward;[15] or they send patients to other hospitals. In obstetrical cases the options are limited; emergent babies cannot be asked to wait, and state laws restrict the use of other facilities for obstetrical cases because of the danger of infection.[16] When an obstetrical unit is full, patients must be crowded into existing facilities, sent to another institution, or other patients must be dismissed early.

As depicted in Figure 1, the average penalty cost per case falls as the size of the unit increases because with larger size the number of patients requiring special handling falls.[17] The three costs components, hospital, travel and penalty costs, together constitute total cost; it is this sum which the community wishes to minimize.

Assuming that births are like horse kicks, that is, described by the Poisson distribution, and that the length of hospital stay is described by a series of hyperexponential distributions,[18] it is possible to apply a simple queuing theory model to the problem at hand. The model, known as the Erlang loss formula,[19] determines for systems with these

[15] Newer facilities are being built with "swing" beds; that is, beds so situated they can be utilized on more than one service.

[16] In some states, hospitals are allowed to put certain types of nonobstetrical cases in obstetrical beds, but not the reverse.

[17] The number requiring special handling is a function of the coefficient of variation and, therefore, falls as units become larger.

[18] Both assumptions have been tested against actual admissions and length of stay data for the Chicago region. While the Poisson arrival assumption is not rejected, a better model seems to be a Poisson process with a shifting parameter; there is a monthly seasonal, and there appears to be daily variation as well. I have allowed for the seasonal movement in the calculation of facility requirements. The length of stay distribution is well approximated by a k-Erlang distribution of the third degree. In addition, I have assumed that all obstetrical units are producing the same product: maternal care. This assumption could be modified to take into account differing qualities of service.

[19] The Erlang loss formula is:

$$P_{N \to K} = \frac{\frac{1}{N}!(x)^N}{1 + x + \frac{1}{2}!(x)^2 + \cdots + \frac{1}{K}!(x)^K}.$$

The probability $(P_N = {}_K)$ of the unit being full when no queuing is permitted is given by the above formula; N denotes the number of customers, K the number of serving channels and X the mean census. For those not familiar with queuing theory, the conditions under which this model is applicable are discussed in D. Cox and W. Smith [10, p. 47] and P. Morse [18, pp. 31 and 45]. Tables of the Erlang loss function have been published only for small values of the parameters. Computer routines are available from the authors which will calculate for any system the probability of being full or the number of servers needed for a given probability.

characteristics and known parameters the percentage of patients[20] that a hospital with a given number of beds will be forced to handle in a special way because all regular beds are occupied.

To illustrate the principles involved, we have constructed a hypothetical system but one in which the number of births is equivalent to that in the Chicago region.[21] The results reported in Table 1 show how the number of required beds,[22] cases requiring special handling, and travel times would change for systems composed of various numbers of obstetrical units and operating under different rules. The calculations presume that all obstetrical units are of equal size and that the units are optimally distributed over the space served.[23] Column (1) shows total bed requirements for a system of independent hospitals each sufficiently large to reduce the probability of being full to one in 10,000. Column (2) shows the system bed requirements[24] when the

TABLE 1

OBSTETRICAL SYSTEM CHARACTERISTICS* BY NUMBER OF UNITS

| Number of Units | Beds | | | | Number of Patients Requiring Special Handling | | Total Travel Time |
| | Probability | | Penalty | Cost | | | |
	.0001 (1)	.001 (2)	$500 (3)	$250 (4)	$500 (5)	$250 (6)	(7) Hours (000)
45	3,240	2,989	2,623	2,430	1,963	5,134	1,422.7
55	3,383	3,102	2,679	2,464	2,198	5,715	1,295.2
65	3,515	3,209	2,729	2,495	2,418	6,264	1,187.3
75	3,635	3,305	2,776	2,523	2,622	6,782	1,108.8
85	3,759	3,407	2,821	2,549	2,810	7,269	1,040.0
95	3,868	3,493	2,859	2,571	2,989	7,724	981.2
105	3,974	3,577	2,896	2,591	3,160	8,164	932.1
115	4,081	3,664	2,930	2,610	3,324	8,588	892.9
125	4,173	3,736	2,963	2,626	3,485	9,012	858.6

* The expected number of deliveries in this system is 157,000 per year.

[20] In this model the percentage of patients turned away and the percentage of time the facility is full are equivalent.

[21] The area covered, as defined by the Hospital Planning Council of Chicago, includes Cook, DuPage, Lake, McHenry, Kane, and Will Counties in Illinois and Lake and Porter counties in Indiana. In this area in 1963, 157,000 maternity cases were handled in the obstetrical units of 95 hospitals. The units contained a total of 3,229 beds.

[22] Beds are the only available measure of facilities, and throughout this study beds are used as the measure of size. A decrease in beds, therefore, represents a decrease both in staff and facilities, though the change in other inputs may not be proportional to the change in beds. In a more sophisticated model of obstetrical care at least two types of facilities should be considered: delivery facilities and bed facilities. Either type of facility can become a bottleneck. The appropriate relationship between the two types of facilities has been studied by A. Barr [2], and J. Thompson and R. Fetter [21].

[23] Actually we have avoided the problem of spatial distribution by assuming a uniform density of patients over the entire area. For a discussion of locational questions, see W. Carr [4].

[24] For example, in a system of 45 units each with 72 beds the probability of any one being full is one in 10,000. Total beds in the system would be 3,240.

target probability of being full is one in 1,000.[25] It is clear that more
beds are needed if the probability is set low or if the system is com-
posed of many small units. Columns (3) and (4) give system bed re-
quirements when the penalty cost associated with cases requiring spe-
cial handling is either assumed to be $500 or $250 per case. The deri-
vation of these results is explained below.

Few hospitals do the type of cost accounting which permits obstetri-
cal costs to be separated from other costs. The best data we could find
to use in estimating obstetrical cost functions were two annual cross-
sections, 1964 and 1965, of 34 Connecticut hospitals.[26] The two regres-
sions shown in Table 2 are linear and have the total costs of the obstet-

TABLE 2

TOTAL OBSTETRICAL COST REGRESSIONS*

Date	R^2	Statistic	Intercept	Independent Variables		
				Beds	Patients	Length of Stay
1964	0.97	Coefficient	−219,200	5,135.3	135.3	39,910
		T Value	−3.0	3.1	4.9	2.2
1965	0.97	Coefficient	−230,100	5,613.1	142.8	41,390
		T Value	−3.0	2.4	3.8	2.1

* There were 34 observations for each regression.

rical units as the dependent variable. There is high collinearity be-
tween two of the independent variables, the number of beds and the
number of patients, but the coefficients of· these variables are
significant and remain stable between the two years. The coefficient of
the variable "beds" in the 1965 regressions has been used as the esti-
mate of savings from reducing hospital beds by one, the number of pa-
tients and the average length of stay held constant.[27] As it stands the
cost function is probably not specified correctly; significant variables
have almost certainly been excluded. The two most obvious omitted
variables are wage rates and an allowance for differences in the quality
and type of output. No good wage series was available and the one we
constructed did not add significantly to the explained variance. Cor-
recting for quality differences is the bane of every investigator of hos-
pital cost functions.[28] Available data do not adequately measure out-

[25] If the probability of being full is one in 10,000, special arrangements would have to be
made on average for 15.7 patients per year, and in the one in 1,000 case for 157 patients per
year.
[26] These hospitals do attempt to allocate all joint costs to particular services.
[27] In recent years the length of stay for maternity cases has been getting shorter. The cost
functions indicate that a further reduction, say from the present mean in Chicago of 4.5 to
4.0 days, would result in about a 6 percent fall in the cost of caring for maternity patients.
[28] A partial listing of recent studies of hospital costs includes: P. Feldstein [13], M. Ing-
bar [14], W. Carr and P. Feldstein [5], M. Feldstein [12], K. Ro [19], R. Berry [3], and
G. Wirick [23]. For a review of some of these studies and several others, see J. Lave [15].

put, not only because of quality differences, but also because some hospitals produce ancillary services of training and research. In the data employed, the latter problem was presumably handled through the cost accounting procedures and the quality problem should not have been important since complicated deliveries constitute only a small percentage of the total. Yet the negative constant terms in the regressions imply that there are diseconomies of scale, that significant variables have been left out of the model, or that the functional form is not linear.[29] Other investigators have concluded that once adequate correction has been made for output differences, there are either scale economies (Berry [3], R. Coughlin [8], Ro [19], P. Feldstein [13] and W. J. Carr and P. Feldstein [5]) or constant (Ingbar [14] and M. Feldstein [12]) returns to scale in hospitals. Hospital administrators with whom we talked were also convinced there would be no diseconomies as scale size was increased. Since the omitted variables are probably positively correlated with size, their exclusion has most probably resulted in upwardly biased estimates of the coefficients on patients or beds, or both.[30]

It is difficult to determine the appropriate penalty to assign when facilities are full. In most obstetrical units one or two extra patients can probably be accommodated by crowding at rather small cost; severe crowding or dismissing patients early probably involve higher costs; the highest costs under present institutional arrangements probably occur when patients are transferred to another hospital. Delivery dates are, of course, predictable in advance with considerable accuracy, which makes possible the scheduling of admissions. Thus sending a patient to another hospital need not be an emergency procedure, though it does involve some inconvenience to both patient and doctor. In the case of transfers, the important question is whether the doctor has a staff appointment at another institution where he can deliver the mother conveniently; if he does not, the costs of the hospital being full may be quite high. But as only a few patients need be transferred and most doctors in metropolitan areas have appointments at least at two hospitals,[31] it should be possible, even with present institutional arrangements, to find cases which could be transferred at reasonable cost. For present purposes we have assumed two values for penalty

---

[29] The scatter diagrams do suggest that if the largest hospitals were eliminated, the intercept would be positive. While we did not try this, we did run some regressions in forms other than linear.

[30] M. Feldstein's estimates [12, Table 5, p. 198] of obstetrical costs in England show ratios between the "fixed" costs associated with beds and the "variable" costs associated with patients that are very similar to those reported in this study.

[31] Our data showed that half the obstetricians in the Chicago region delivered women in more than one hospital in the single month of Feb., 1965. Had the data been available for a longer period, it would probably have revealed that more than half the obstetricians in Chicago are connected with several hospitals.

costs, $500 and $250 per case requiring special handling.[32] Community
costs are minimized when the marginal savings from reducing beds are
equated to the increased costs associated with patients requiring spe-
cial handling.[33] Columns (3) and (4) of Table 1 give the system bed
requirements resulting from these calculations. Columns (5) and (6)
indicate the numbers of patients who could not be handled in a routine
way were hospitals of the size specified.

By employing the above techniques the optimal size of a unit can be
determined, given the number of units. To calculate the optimal num-
ber of units in a system, travel costs and economies of scale must also
be considered. Travel time estimates were derived from data on the
time and distance between home and hospital for obstetrical patients
in the Chicago region.[34] Estimates were then made of how travel time
would change as the number of units increased or decreased from pres-
ent levels.[35] Coughlin et al. [9] found that in Philadelphia inpatients
accounted for only 5.9 percent of the total trip miles involved with
hospitalization. Visitors to inpatients accounted for over half the travel
distance; employees, outpatients and physicians made the remainder
of trips. Assuming Coughlin's figures hold generally, a reduction in
units that adds one mile to the average travel distance increases total
travel roughly thirty-four miles per case.

Using the information available and a few additional assumptions,
the physical magnitudes, beds, transfers and hours of travel given in
Table 1, can be turned into costs estimates; values can then be found
which will minimize the total costs to the community of obstetrical ser-
vice. To carry out these calculations requires temerity; but the results
are reported with considerable diffidence. Assuming there are no scale
economies, that penalty costs are between $250 and $500 per case re-
quiring special handling and that the value of travel time is between
$1.55 and $2.80 per hour,[36] then the optimal system for the Chicago
region would have between 80 and 125 units and 2,534 and 2,963 beds.
The lower values cited are optimal if penalty costs are $250 per case
and travel time worth $1.55 per hour; the higher figure holds if the
penalty cost is $500 with travel time worth $2.80 per hour.

[32] We are currently engaged in further research on the question of penalty costs. Instead
of a flat rate it would appear to be more appropriate to assume penalty costs to be an in-
creasing function of the number of cases which cannot be handled in routine fashion.
[33] Some of the costs may be internal to the hospital; most, however, will fall upon pa-
tients and doctors.
[34] The available data were for a sample of Chicago obstetrical patients for the month of
Feb., 1965.
[35] It was presumed that the time in travel was proportionate to the distance traveled and
that the latter was proportional to the square root of the ratio of 95 (the present number
of obstetrical units) to the hypothecated number. No correction was made for the additional
travel time of those who could not be handled in the hospital of their choice but the effect of
this on total travel time would be small.
[36] The source of these estimates of the value of travel time is H. Mohring [17]. Travel costs,
other than time, are not considered.

In addition to the above hypothetical model, we have examined the situation for each obstetrical unit in the Chicago region. First, it is clear that due to shifts in population, the existing beds are far from optimally distributed and the probability that some hospitals will be full is much higher than in others.[37] With the existing distribution, the expected number of patients who must be handled in some special way is over 2,000 per year, but if the same number of beds were optimally distributed this number could be reduced.

We have calculated bed needs for the existing units on the assumption that the cost of special handling is $250 and $500 per case. With the existing pattern of demand and maintaining the same number of units, it would be possible to reduce facilities from the existing 3,002 beds[38] to 2,656 if penalty costs were $500 per case, and to 2,412 if the cost were $250.[39] This would raise overall occupancy levels in obstetric units from the present 60 percent to 67 percent in the one case and 75 percent in the other. Most of the contraction would come in the older central city units, while some expansion would be called for in some of the suburban areas. Arrangements of the type suggested would reduce costs of obstetrical service to the Chicago community by about 4 percent of total outlays were the appropriate penalty figure $500 and 8 percent if $250 were correct; that is, between $1.5 million and $3 million.[40]

Policy recommendations usually call for the amalgamation of smaller obstetrical units to achieve higher occupancy.[41] This report has suggested that physical merging is unnecessary, especially if the community is prepared to transfer patients when one unit is full. In essence the choice lies between making all patients travel to larger units or with an integrated system having the marginal patients travel when one unit is full.

Short-term general hospitals constitute a large industry with outlays of more than $9 billion per year, so that even a small percentage savings is large in absolute amount. Our preliminary research indicates that the same procedures applied on other services would produce less of a savings in facilities but that the value per unit reduction would be greater. But at this stage of the study, it is somewhat dangerous to

[37] For some hospitals in the Chicago region the probability of being full is very small; for others it is as large as one in 10. As might be expected, those with small probabilities, in essence unneeded beds, are the older central city hospitals.

[38] Lack of data made it impossible to include in the calculations 10 of the existing 95 units.

[39] In all of these calculations an allowance has been made for the seasonal variations in birth patterns.

[40] The estimates are adjusted for penalty costs. In-hospital savings would be 5 percent and 9 percent in the $500 case and $250 case, respectively.

[41] See, for example, Hospital Review and Planning Council [26]. The study named assumes that hospitals will act independently rather than cooperate with their peak load problems; moreover, the study is interested in hospital not community costs of care.

generalize our results either to other geographical areas or to other services.

Regional planning for hospitals involves more problems than those discussed in this paper. For regional planning to be successful, accurate forecasts are needed not only of the overall demand but of the demand for particular services at various hospitals. Furthermore, we have taken the in-hospital cost curve as given; others have studied the possibility of changing the parameters of the costs function through different organizational arrangements. (See Coughlin [8], Schneider [20], and J. Young [24].) A third aspect of regional planning neglected in this paper is the quality of care. The recent President's Commission [27] suggested that different institutional arrangements among hospitals within a region could improve the quality of care.

Improved techniques for handling peak loads in hospitals can reduce the costs of care. As these arrangements will reduce staff needs, they will help relieve the nursing shortage. Another implication of the analysis presented, and this applies to all categories of service, is that hospital systems will be more efficient if particular institutions are not restricted to serving particular populations, such as veterans or charity patients or Catholics, because to provide for the peaks in demand for each group separately requires more beds than if the groups are not differentiated. This means, for example, that it may be more efficient for governments to pay for the hospitalization of veterans or indigents in regular hospitals than to provide service directly.

REFERENCES

1. A. D. Airth and D. J. Newell, *The Demand for Hospital Beds* (Newcastle-upon-Tyne, 1962).
2. A. Barr, "Computer Simulation of a Maternity Hospital," *British Hospital J. and Soc. Serv. Rev.*, Mar., 1965, pp. 507-10.
3. R. Berry, "Competition and Efficiency in the Market for Hospital Services: The Structure of the American Hospital Industry" (unpublished doctoral dissertation, Harvard Univ., 1965.)
4. W. J. Carr, "Notes on Area Wide Planning" (mimeo.), 1966.
5. W. J. Carr and P. J. Feldstein, "The Relationship of Cost to Hospital Size" (mimeo.), 1966.
6. J. H. Cavanaugh, "The Rise of the Area Wide Planning Agency: A Survey Report," *Hospitals*, Aug., 1965, pp. 52-56.
7. R. Coase, "The Problem of Social Cost," *J. of Law and Econ.*, 1960, pp. 1-44.
8. R. E. Coughlin, "Hospital Complex Analysis: An Approach to Analysis for Planning a Metropolitan System of Service Facilities" (unpublished doctoral dissertation, Univ. of Pennsylvania, 1965).
9. R. E. Coughlin, W. Isard, J. B. Schneider, "The Activity Structure and Transportation Requirements of a Major University Hospital," *Discussion Paper Series*, No. 4, Reg. Sci. Res. Inst., Philadelphia, 1964.
10. D. R. Cox and W. L. Smith, *Queues* (London, 1961).
11. P. de Vise, *Hospital Study Districts for Metropolitan Chicago,* Hospital Planning Council for Metropolitan Chicago, Technical Report No. 2, Chicago, 1966.
12. M. Feldstein, "Economic Analysis for Health Service Efficiency" (unpublished doctoral dissertation, Nuffield College, Oxford, 1966).
13. P. Feldstein, *An Empirical Investigation of the Marginal Costs of Hospital Service* (Chicago, 1961).

14. M. L. Ingbar, "A Statistical Study of Differences in Hospital Costs: Cost Functions for 72 Massachusetts Hospitals" (mimeo.) (Grad. Sch. of Pub. Admin., Harvard Univ., 1964).

15. J. R. Lave, "A Review of the Methods Used to Study Hospital Cost," *Inquiry*, May, 1966, pp. 57-81.

16. M. F. Long, "Efficient Use of Hospitals" in *The Economics of Health and Medical Care* (Ann Arbor, 1964), pp. 211-26.

17. H. Mohring, "Urban Highway Investments" in R. Dorfman, ed., *Measuring Benefits of Government Investments* (Washington, 1965, pp. 231-91.

18. P. M. Morse, *Queues, Inventories and Maintenance* (New York, 1958).

19. K. K. Ro, "A Statistical Study of Factors Affecting the Unit Cost of Short-Term Hospital Care" (unpublished doctoral dissertation, Yale Univ., 1966).

20. J. B. Schneider, "Planning the Growth of a Metropolitan System of the Public Service Facilities: The Case of Short-Term General Hospitals" (unpublished doctoral dissertation, Univ. of Pennsylvania, 1966).

21. J. D. Thompson and R. B. Fetter, C. S. McIntosh and R. J. Pelletier, "Predicting Requirements for Maternity Facilities," *Hospitals*, Feb., 1963, pp. 45-49, 132.

22. B. Weisbrod, "Some Problems of Pricing and Resource Allocation in a Non-profit Industry—The Hospitals," *J. of Bus.*, 1965, pp. 18-28.

23. G. C. Wirick, "Hospital Cost Study: Econometric Analysis" (mimeo.) (Ann Arbor, 1965).

24. J. P. Young, "A Queuing Theory Approach to the Control of Hospital Inpatient Census" (unpublished doctoral dissertation, Johns Hopkins Univ., 1962).

25. Joint Committee of the American Hospital Association and Public Health Service, *Area Wide Planning for Hospitals and Related Facilities* (Washington, 1961).

26. Hospital Review and Planning Council, *Guidelines and Recommendations for the Planning and Use of Obstetrical Facilities in Southern New York* (New York, 1966).

27. The President's Commission on Heart Disease, Cancer and Stroke, *A National Program to Conquer Heart Disease, Cancer and Stroke* (Washington, 1965).

THE ECONOMIC EFFECTS OF MALARIA ERADICATION*

By ROBIN BARLOW
University of Michigan

In this paper a method is presented for measuring the economic effects of malaria eradication, in particular the effects of eradication on per capita income in the long run. The method, which involves specifying a fairly detailed model of the entire economy, has been applied to the case of Ceylon, and some preliminary results are presented on what the course of Ceylonese per capita income would have been after 1947 if the successful eradication campaign of that year had not been undertaken.

These inquiries may be of interest to the economic historians, as providing an understanding of the far-reaching consequences of eradication programs which have been executed in the past. The inquiries may be relevant also to the decision-maker in government who must choose between alternative expenditure proposals in the context of a budget constraint. Considerations of the economic payoff of the alternative projects may properly influence the decisions which are made in those circumstances.

At the outset it should be stressed that in drawing attention to the economic consequences of disease eradication, the economist does not mean to belittle the noneconomic consequences. Many public health officials seem to feel that when the economist points to the possible economic losses resulting from eradication—such as those implied by a population explosion—he is callously arguing that people should be denied the benefits of modern medicine. The fact is that an eradication program, like all other government programs, ought to be judged by the contribution which it makes toward the several national goals. In most countries the growth of per capita income is a national goal which is accorded a fairly high priority, and therefore the effect of eradication on per capita income is a relevant consideration. It may of course be rational to adopt an eradication program even when it is known that per capita income will probably fall as a result. This may be so when the program makes large contributions to other goals, such as the diminution of suffering.

Malaria and Its Eradication[1]

Malaria is a parasitic disease transmitted by certain species of the anopheline mosquito. The parasites invade red blood cells and cause them to rupture synchronously, thus producing the attacks of chills, fever, and sweating associated with the disease. These attacks cause temporary disability and are sometimes fatal. Those victims who survive the febrile attacks are left in a weakened state because of the massive destruction of their blood cells, and hence are more susceptible to death or disability from other diseases. Their debilitation reduces work efficiency. Eventually the blood cells are replaced and the body may develop some temporary immunity against a further infection. In many areas of the world, malaria has been endemic; in others it has appeared spasmodically or in epidemic form because of such factors as extreme instability in the size of the mosquito population. The annual number of cases on a worldwide basis has been reckoned in the hundreds of millions.

After the discovery some seventy years ago of the means whereby the disease was transmitted, the methods of control or eradication have multiplied rapidly. No universal panacea has been found, and one of the chief tasks in an eradication campaign is to choose from among the numerous possibilities that particular set of measures which is best suited to the local environment. The measures available include the use of mosquito netting, the location of dwellings at an adequate distance from the mosquitoes' breeding places, the use of insect repellents, the administration of drugs such as quinine which reduce or eliminate the parasite population in the human body, the installation of new drainage systems, or the removal of vegetation to render the mosquitoes' breeding places unusable, larvicidal techniques such as the oiling of ponds and streams and (often the most effective of all) residual spraying of house interiors with DDT or other insecticides.

Before the second World War attempts at controlling malaria were mostly localized and unambitious. In the years immediately following the war several campaigns were organized which made use of the new insecticides and antimalarial drugs then becoming available and in many cases virtually complete eradication was achieved over extensive areas. The possibility therefore emerged that the disease could be rendered totally extinct, and in 1955 the World Health Organization formally established worldwide eradication as one of its goals. In the pursuit of that goal the World Health Organization and its affiliates

[1] For a full treatment of these matters, see one of the standard works on malariology such as Emilio Pampana, *A Textbook of Malaria Eradication* (Oxford Univ. Press: London, 1963) or Paul F. Russell *et al.*, *Practical Malariology*, 2nd ed. (Oxford Univ. Press: London, 1963).

have provided substantial technical and financial assistance to the national eradication campaigns. The proceedings have assumed a note of urgency as more cases of resistance by mosquitoes and parasites have been disclosed, and the possibility now exists that worldwide eradication will be indefinitely postponed unless it is accomplished quickly, before the resistant strains become established.

By 1966 the eradication campaigns have proceeded to the point where the population of territories freed from malaria exceeds the population of territories where malaria still prevails. Since a campaign cannot succeed without money and some organizational resources, it is naturally the poorest and most backward areas of the world which still suffer from the disease. These areas include most of tropical Africa and parts of Central and South America, the Middle East, the Indian subcontinent, and Southeast Asia.

Malaria eradication is a half-finished task. In some countries the disease was virtually eradicated two decades ago or more, and these cases could provide abundant evidence on the economic and social consequences of eradication. Other countries are still malarial and are in a position to benefit from the experiences of the first group when they come to decide upon the timing, scale, and methods of their own eradication programs.

A Classification of the Economic Effects of Eradication

In the public health literature there is a consensus that malaria eradication is economically · beneficial. Occasional misgivings are expressed about the implications of a population explosion; but even when these possible disadvantages are admitted, it is invariably said that they are far outweighed by the economic gains from eradication. These sanguine views have not been based on sound evidence. There have been several casual estimates purporting to demonstrate the economic advantages of eradication.[2] Some of these estimates have involved measuring the additional number of man-days of work which eradication has made possible in a particular country through reducing mortality and sickness, and then multiplying this number by an average daily wage. Other estimates have compared the costs of the eradication campaign itself to the extra hospitalization expenses which would have been incurred if malaria had been allowed to persist. All of the estimates have been incomplete: they have failed to take into account the multiplicity of economic effects.

In providing a more complete analysis, we can begin by stipulating that per capita national income is the economic variable whose fate

[2] For an account which describes many of these estimates, see C.-E. A. Winslow, *The Cost of Sickness and the Price of Health* (W. H. O. Monograph Series, No. 7: Geneva, 1951).

concerns us. Eradication could affect income per capita by changing either the number of heads or the level of income. The level of income or output could be affected by changes in the quantity and quality of labor inputs, the quantity and quality of capital or nonhuman inputs, and the manner in which these inputs are combined. Hence there are four categories of effects to consider. In our discussion of these effects we shall not dwell on the problems of measurement. These problems will be aired in the next section, where a model is outlined which incorporates the effects discussed below.

Effects on Population Size. Other things being equal, a rise in the rate of population growth will reduce per capita income. There is evidence that malaria eradication produces this result both by lowering death rates and by raising birth rates. The fall in death rates occurs, not only because of a reduction in deaths directly attributable to malaria, but also because the population acquires a greater resistance to other diseases.

The rise in birth rates which has often been observed to follow eradication can probably be explained by the fact that pregnant women attacked by malaria are more liable to suffer miscarriages. There is also the possibility that the rate of conception is lower in malarial conditions because of the reduced vitality of both men and women.

Effects on Labor Inputs. Other things being equal, a rise in the quantity or quality of labor inputs will cause per capita income to rise. Eradication can affect labor inputs by reducing mortality, morbidity, and debility.[3] The reduction in mortality will bring about an immediate increase in the quantity of labor inputs in cases where the fall in death rates occurs among those of working age. A fall in death rates among those in younger age groups will cause an expansion of the labor force after a lag.

A reduction in morbidity, which is defined here as sickness sufficient to cause absence from work, can bring about an immediate increase in the quantity of labor inputs. A reduction in debility, which is defined as affecting the worker's productivity on the job, can cause an increase in the quality of those inputs. A worker freed of the debilitating effects of the disease can improve his performance both physically and mentally; he may complete a given manual task in a shorter time and also undertake activities more imaginative and ambitious than before.

The economic payoff to disease eradication depends not only on the relative incidence of the disease in the labor force and the rest of the population. It depends also on what segments of the labor force are benefited by eradication. In the case of malaria we are dealing with a

---

[3] These distinctions are discussed in Selma J. Mushkin, "Health as an Investment," *J.P.E.,* Oct., 1962, pp. 129-57.

disease whose greatest impact is upon the low-income, low-productivity segment of the labor force, and the payoff to eradication is therefore less than would be the case with a disease which involved the same number of victims but which was concentrated among the most productive segment. Malaria is a low-income disease mainly because it is avoidable. Persons with higher incomes and higher levels of education are in a stronger position to take advantage of that characteristic: being better informed, they are more likely to avoid exposure, for example by staying away from mosquito-infested places during darkness; being wealthier, they are better able to afford preventive or curative measures, such as the use of quinine or mosquito netting. From the economic viewpoint, malaria therefore contrasts sharply with such diseases as cancer or heart disease, where the opportunities for prophylaxis are distinctly limited and where the incidence tends to be the greatest among the most productive members of the labor force.

Effects on Capital Inputs. Other things being equal, the higher the rate of capital formation, the more rapid will be the growth of per capita income in the future. We must therefore examine the effects of disease eradication on the division of total expenditures between consumption goods and capital goods in both the private and public sectors.

In the private sector the larger population which results from eradication is likely to lower the rate of saving (and hence the rate of capital formation) attainable from a given level of disposable income. Moreover, when the population is growing more rapidly, what limited private saving does occur may tend to be invested in housing, which is a relatively unproductive form of capital stock. There will be some offsetting effects on the disposition of private income. After a government program of eradication, private expenditures for the prevention or treatment of the disease become unnecessary. The funds thus released may be devoted in part to saving and capital formation.

In the public sector the allocation of funds between consumption and investment is governed by political priorities which vary from country to country, and it is less easy to generalize about the effects of eradication. For our purposes, public investment should be defined as outlays which add to the stock of "productive" physical capital (like dams, roads, or government factories); all other resource-using governmental expenditures should be defined as public consumption, even when they provide long-lasting real assets (like police stations, schools, or sports stadia). One fairly common situation is for the highest budgetary priorities to be assigned to various traditional forms of public consumption, with any residual revenues being made available for new projects of physical investment. Other things being equal, the larger

population resulting from eradication may necessitate an expansion of the traditional services and thus leave a smaller residuum for public capital formation. In this setting the direct expenses of the antimalaria campaign itself cause an equivalent reduction in public investment. At the same time the lessened morbidity may permit the government to reduce its expenditures for medical care and funds would therefore be released for public investment as defined.

It should be recognized that many of the high-priority consumption expenditures of government promote the growth of income through improving the quality of the labor force even though they do not add to the stock of physical capital as defined. It would seem advisable to pay particular attention to the educational sector in assessing the economic effects of malaria eradication. In most countries of the malarial zone, education is a large-scale high-priority government service; the rapid increase which eradication produces in the population of school age therefore means that a large quantity of extra funds must be committed to education. These expenditures create a more productive labor force, but only after a substantial lag. If eradication did not occur, the funds in question could by assumption be invested to bring an immediate payoff.

Other Effects on Output. The level of output depends not only on the quantity and quality of human and nonhuman inputs but also on the manner in which those inputs are combined. It is possible for eradication to increase output by inducing a change in input combinations. The effect in question has been recognized and indeed exaggerated in the public health literature. It is said that malaria eradication permits the exploitation of new territories which previously were shunned because of the threat of disease. Very often the gross value of the output from the newly exploited districts is cited as a measure of the benefits from eradication. It is not acknowledged that the capital equipment and labor used in these districts presumably had some positive opportunity cost, and that the increase in output from the previously malarial districts is obtained at the price of reduced output elsewhere.

Nevertheless, the phenomena in question may involve net gains for the economy. These gains are perhaps best analyzed as stemming from a spatial reallocation of resources which is induced by a localized decline in the disutility of work. Eradication lowers the disutility of work in the previously malarial districts; labor accordingly tends to migrate to those districts; the shift in labor raises the marginal product of capital in those districts relative to its marginal product elsewhere, and in the long run a migration of capital will accompany the migration of labor. If, as seems likely, the marginal product of land in the malarial districts before eradication was higher than its marginal

product elsewhere, eradication will have contributed to an expansion of output by causing labor and capital to be relocated in the districts where the marginal product of land was relatively high.

A Model for Measuring the Economic Effects of Eradication

If we wish to measure the impact of eradication on per capita income while allowing for all the effects discussed in the preceding section, it is clear that we must make use of a fairly detailed model of the entire economy. The relationships which appear to be relevant should be specified in the form of equations. The model should then be used to provide two simulations of the course of per capita income through a period of n years, the first simulation being based on the assumption that eradication occurs at the start of the period, and the second simulation being based on the assumption that no eradication campaign is undertaken during the period. The method could be used retrospectively, measuring the consequences of a completed program, or it could be used to predict the outcome of a proposed program. One particularly successful application of the method is to be seen in the measurement by Coale and Hoover of the economic gains from birth control in India.[4]

In specifying a model of the entire economy, this approach is somewhat more elaborate than that typically used in "benefit-cost" analyses of health programs. The typical analysis in this field has been of a partial nature, a legitimate approach when the program in question causes nothing more than marginal changes in the structure of the economy. But malaria eradication can cause quite marked changes. By using Newman's results concerning the effects of eradication on birth and death rates,[5] it can be estimated that the population of Ceylon in 1977, thirty years after eradication, will be 16.0 millions, whereas in the absence of eradication the figure would be only 12.6 millions. It is estimated that in 1977, 27 percent of the population will be in the labor force; without eradication the figure would be 29 percent. In 1977, the education budget in Ceylon will be about 700 million rupees; the figure would be about 500 million rupees in the absence of eradication. The methods of partial analysis, which would essentially ignore the indirect repercussions of eradication, would probably yield misleading results in this situation.

A model has therefore been designed which allows for the major economic effects, direct and indirect, of malaria eradication. With some modifications the model would also be useful for evaluating any public

[4] Ansley J. Coale and Edgar M. Hoover, *Population Growth and Economic Development in Low-Income Countries* (Princeton Univ. Press, 1958).
[5] Peter K. Newman, *Malaria Eradication and Population Growth, With Special Reference to Ceylon and British Guiana* (Bur. of Pub. Health Econ., Univ. of Michigan, 1965).

health program which had important effects on demographic structure and labor supply, such as programs of birth control, programs causing substantial reductions in infant mortality, or the eradication of widespread debilitating diseases like schistosomiasis. The model may even be useful for evaluating nonmedical programs such as education.

The model has been designed with the case of Ceylon in mind, and is being used to estimate the effect of eradication on Ceylonese per capita income during the thirty-year period following the successful campaign of 1947, a period of this duration having been chosen as allowing for the adequate operation of all lagged effects. If other investigators wished to apply the method to other countries, they might find it desirable to amend one or two of the equations which reflect structural peculiarities of the Ceylonese economy.

In the first part of the model, age-specific birth and death rates have been used to generate the age-sex composition of the population year by year. No attempt has been made to specify relationships between the vital rates and the various social and economic variables which appear elsewhere in the model: in a more sophisticated system it would be possible to incorporate, for example, the presumably negative relationship between birth rates and the level of educational attainment among females, or the presumably negative relationship between death rates and lagged per capita income. It is assumed that the vital rates are affected by eradication to the extent estimated by Newman, but no other formal relationships involving the vital rates have been specified. From Newman's analysis it appears that eradication had marked effects on fertility and infant mortality and therefore lowered the average age of the population. It can be estimated that in 1977, 50 percent of the Ceylonese population will be aged under twenty; without eradication the figure would be 45 percent. From the population data thus derived, estimates have been made of the population of "equivalent consumers." The index of economic performance can then take the form of income per equivalent consumer instead of income per capita, which is a somewhat less refined measure.

The main part of the model is devoted to the determination of income, defined as real gross national income. The relevant equations are listed in the Appendix to this paper. There is not the space here to explain the equations fully, and we must be content with mentioning a few of their salient features, along with a brief indication of some of the measurement difficulties encountered in estimating the equations for the case of Ceylon:

1. The production function is of the Cobb-Douglas form, a form chosen because of its relative simplicity and because the terms appearing in the function are readily measurable. The function differs, how-

ever, from the simplest Cobb-Douglas formulation in two main respects: (a) the labor input is disaggregated into two components, skilled and unskilled, skilled workers being defined as those with some secondary education; (b) each of the three inputs—skilled labor, unskilled labor, and capital—is characterized by a quality index. The purpose of disaggregating the labor input is to allow for the fact that malaria eradication has a differential impact on components of the labor force which differ greatly in their productivity. The introduction of the three quality indexes allows for various other consequences of eradication.

2. The indexes of skilled and unskilled labor quality are increased when malaria is eradicated, because debilitation is thereby reduced. The measurement difficulties here are acute, and an empirical investigation of the economics of debilitation must surely have high priority among future research projects in the area of medical economics.[6] In the present study the effect of eradication on the indexes of labor quality was estimated on the basis of opinions expressed informally by malariologists. Additional simulations are being performed on the assumption that the effects are much more powerful than initially supposed, the objective being to see whether the effects of eradication on income per equivalent consumer are highly sensitive to the extent of debilitation.

3. The index of the quality of the capital stock depends on the relative importance of three components of each year's investment: imported capital goods (assumed to have a high quality or productivity), housing construction (low quality), and other investment (medium quality). The larger population occasioned by eradication causes more investment to take the low-quality form of housing construction, although only after a considerable lag. Imports of capital goods will be subject to a balance-of-payments constraint. In the case of Ceylon, the foreign exchange receipts needed for purchasing imports are assumed to be independent of eradication. In particular, eradication is assumed to have no effect on exports: the acreage suitable for the major Ceylonese export crops (tea, rubber, and coconuts) lies mostly outside the previously malarial zone, and export production has not been limited to an important extent by labor shortages. In this context, eradication may generate a larger volume of investment, but that volume is likely to be of inferior quality because it will contain a relatively small proportion of imports.

4. There is a fourth quality index in the production function, an

---

[6] For several suggestions as to how the effects of debilitation could be measured, see Mushkin, *loc. cit.*, pp. 141-42.

ndex of allocative quality. Eradication can affect the level of this ndex by inducing a spatial reallocation of resources, as discussed bove. Guesses are unavoidable in estimating the values of this vari- ble, the only relevant information available in the case of Ceylon)eing that which describes the shift in the labor force between pre- iously malarial and previously nonmalarial districts.

5. The exponents in the production function are the elasticities of)utput with respect to skilled labor, unskilled labor, and capital, re- pectively. If, as seems reasonable, the sum of the three elasticities is :onstrained to equal unity, we can conclude on the basis of the Euler heorem that their values are approximated by the shares of the re- pective factors in total income. Data on the capital share in Ceylon ire not available, and in the present study, simulations are to be)ased on alternative capital-elasticity values of 0.5 and 0.3 (the latter)eing a normal value for Western economies). Data on the relative hares of skilled labor and unskilled labor are available from the *Sur-)eys of Consumer Finances* undertaken by the Central Bank of Ceylon n 1953 and 1963. The Central Bank *Surveys* are also invaluable in)roviding data needed elsewhere in the model, such as data on private nedical expenditures and on the proportion of each age-sex group who lave received secondary education.

6. The quantity of labor inputs depends in part on morbidity rates. :he reductions in these rates following the Ceylonese antimalaria cam-)aign can be roughly estimated from data collected by Cullumbine in a lousehold survey of morbidity.[7] Cullumbine's data suggest that eradi- ation caused significantly greater declines in morbidity among the un- killed than among the skilled.

7. The value of the capital stock in the initial year, along with the 'alue of the constant term in the production function, can be obtained)y simultaneous solution of two versions of the production function—)ne version containing the values for one arbitrarily selected year and he other containing the values for some later year—in which these wo terms are the only unknowns. The capital stock grows annually as . result of net investment. Net investment is assumed to be solely de-)endent on the supply of saving, an assumption which is justifiable for)w-income countries although not for high-income countries. Most of he data needed for simulating the growth of saving are to be found in . detailed set of national accounts. In the case of Ceylon these data .re obtainable from the Central Bank's *Annual Reports,* the Census

[7] H. Cullumbine, "A Survey of Disabling Illness in Ceylon," *Bul. of the World Health)rganization,* VII (1952), pp. 405-29, and "The Health of a Tropical People," *The Lancet,* /lay-June, 1953, pp. 1090 ff., 1144 ff., 1193 ff., and 1245 ff.

Department's *Statistical Abstracts,* the United Nations' *Yearbooks of National Accounts Statistics,* and in particular from a recent study by Snodgrass.[8]

8. Total saving consists of private saving, government saving (the excess of tax receipts over public noninvestment expenditures), and foreign saving (the current-account deficit in the balance of payments). Private saving is held to be a function of private disposable income per equivalent consumer. In the determination of government saving, it is assumed that the budgetary priorities are like those discussed above: the public noninvestment expenditures—educational spending, other forms of public consumption, and transfer payments—are dependent primarily on population size, and public investment is essentially a residual. Eradication causes an expansion of public noninvestment expenditures but does not cause a corresponding increase in tax receipts, at least in the case of Ceylon. On the basis of the administrative realities in underdeveloped countries, it is assumed that tax receipts are a function of the size of the foreign trade sector. Since, according to our earlier argument, eradication has done little to increase the size of the foreign trade sector in Ceylon, it has also done little to increase tax receipts. In Snodgrass' words, "caught between . . . rising service bills and the relative inelasticity of government revenues, capital expenditures suffered."[9]

9. Imports of both capital goods and noncapital goods are assumed to be a function of (*a*) demand considerations and (*b*) the size of foreign exchange reserves at the beginning of the year in question. When reserves become seriously depleted, it is common for the government to curtail imports by imposing tariffs or quotas. It was thought unlikely that the relationship between imports and prior reserves would be linear. To provide a more plausible relationship, the square root of the reserves was therefore used in the import equations.

Model Dynamics. It will be clear from the preceding discussion that eradication has numerous positive and negative effects on income per equivalent consumer. These effects do not occur all at once but in a staggered fashion, and the model therefore involves several lags. Some of the more important of these lags are listed in Table 1. On the positive side it is shown that there are some strong effects occurring in the first year after eradication. The lags in the negative effects can be best seen by following the career of the first cohort of "eradication babies"; that is, those babies who would not be alive one year after the eradication campaign if the campaign had not occurred. One should add that the second cohort is appreciably larger than the first: in the second

[8] Donald R. Snodgrass, *Ceylon: An Export Economy in Transition* (Irwin, 1966).
[9] *Op. cit.,* p. 194.

TABLE 1

SELECTED LAGS IN THE EFFECTS OF MALARIA ERADICATION ON
NATIONAL INCOME PER EQUIVALENT CONSUMER

Number of Years after Eradication	Positive Effects	Negative Effects
0		Resources are devoted to eradication campaign
1	Quantity and quality of labor inputs are increased due to reduced mortality morbidity, and debility; private expenditures on medical care are reduced; spatial reallocation of resources begins	Population grows due to reduced mortality
2		Population grows due to increased fertility
6		Part of first cohort of "eradication babies" enters primary school
11		First cohort becomes "equivalent consumers"
15	Part of first cohort enters unskilled component of labor force	Part of first cohort enters secondary school
16		Females in first cohort enter child-bearing age bracket
20	Part of first cohort enters skilled component of labor force	Part of first cohort enters institutions of higher education
21		First cohort enters housing market
26		Females in first cohort enter age bracket where fertility at maximum

year after eradication, not only are infant mortality rates at their new lower levels but birth rates begin to rise.

The Model Applied to Ceylon

The model described in the preceding section and in the Appendix has been used to simulate the course of income per equivalent consumer in Ceylon on the assumption that malaria was not eradicated. Some partial results of the simulation are depicted in Figure 1. At the time of writing, the simulation had proceeded from 1947 through 1955. It can be seen that at the outset eradication had a strong positive impact on income per equivalent consumer. As the years progressed beyond 1947, the gap widened between the with-eradication values and the without-eradication values of income per equivalent consumer, until by
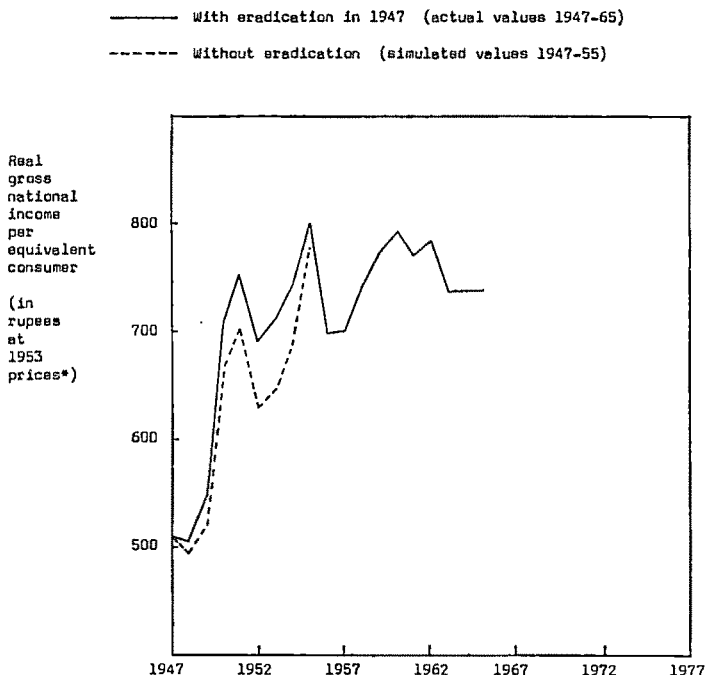
FIGURE 1

THE NET EFFECT OF MALARIA ERADICATION ON NATIONAL INCOME
PER EQUIVALENT CONSUMER IN CEYLON

* In 1953 the official exchange rate was $1.00 = Rs. 4.76.

1953 the income figure with eradication, at Rs. 712, was 10 percent higher than the figure without eradication, at Rs. 646.

Thus in the short run, malaria eradication in Ceylon proved economically beneficial. The explanation for this result can be found most readily by referring to Table 1. Eradication makes an immediate contribution to output by increasing the quantity and quality of labor inputs, primarily through reductions in morbidity and debility, and secondarily through reductions in mortality. In the short run the negative effects of eradication on income per equivalent consumer are weak. The direct expenses of the campaign itself are inconsiderable. During these early years, most of those who are alive because of eradication are infants: not until they are five years old will they begin to impose burdens on the educational sector; not until they are about ten years old will they become fully-fledged "equivalent consumers."

But the negative influences of eradication gather strength as time progresses. Already by 1953 government saving—a major determinant of investment and hence of the future growth of output—was substantially smaller with eradication than without, even though gross nation-

al income was 14 percent higher in the former case than in the latter. By that date the extra population resulting from eradication had begun to make demands on the public sector of such a size that government saving amounted to only Rs. 80 millions; without eradication the figure would have been Rs. 170 millions. Both in 1953 and in 1954 actual net investment fell short of the level it would have attained without eradication. As a consequence, by 1955 income per equivalent consumer, at Rs. 802, was a mere 3 percent higher than the figure without eradication, at Rs. 779. It is likely that future results will show the two curves of Figure 1 crossing beyond 1955.

Further Analysis. Besides extending the two curves of Figure 1 to 1977, future calculations will show whether the results are highly sensitive to the assumptions made about the values of certain critical coefficients which are difficult to measure. Additional simulations will be based on the assumptions (*a*) that the capital-elasticity of output is 0.3 instead of 0.5 as assumed in obtaining the results shown in Figure 1 and (*b*) that the effect of malarial debilitation on labor quality is three times as great as initially supposed.

The main economic disadvantages of eradication are seen to lie in the rapid increase in the population of children resulting from the marked changes in infant mortality and birth rates. A final set of simulations will therefore be performed on the assumption that these disadvantages were avoided in Ceylon through the adoption in 1947 of a twin program of malaria eradication and birth control. It seems likely that the twin program would have made a dramatic contribution to the growth of per capita income.

APPENDIX

Equations for the Determination of Income
An asterisk attached to a variable expressed
in money units (like gross national income)
indicates that the variable is to be measured
at constant prices

Definition of real gross national income

$$GNI_t^* = GDP_t^* + NFI_t^* + TTC_t^* \tag{1}$$

Definition of real net foreign income

$$NFI_t^* = \frac{NFI_t}{pp_t} \tag{2}$$

Definition of terms-of-trade contribution

$$TTC_t^* = \frac{X_t}{pm_t} - X_t^* \tag{3}$$

Production function

$$GDP_t^* = A(qa_t)[(LS_t qs_t)^{eu_t}][(LU_t qu_t)^{eu_t}][(K_t^* qk_t)^{ek_t}]V_t \tag{4}$$

Skilled labor supply

$$LS_t = \sum_i P_{i,t} sk_{i,t}(1 - as_{i,t} - os_{i,t})(1 - ms_{i,t}) \tag{5}$$

Unskilled labor supply

$$LU_t = \sum_i P_{i,t}(1 - sk_{i,t})(1 - au_{i,t} - ou_{i,t})(1 - mu_{i,t}) \tag{6}$$

Definition of capital stock

$$K_t^* = K_{t-1}^* + I_{t-1}^* \tag{7}$$

Investment-saving identity

$$I_t = SP_t + SG_t + SF_t \tag{8}$$

Private saving function

$$SP_t = PC_t\left[\frac{SP_{t-1}}{PC_{t-1}} + sv_t\left(\frac{PDI_t}{PC_t} - \frac{PDI_{t-1}}{PC_{t-1}}\right)\right] \tag{9}$$

Definition of private disposable income

$$PDI_t = GDP_t + NFI_t + TPFATI_t + TR_t - D_t - T_t - MC_t \tag{10}$$

Transfer-payments function

$$TR_t = g_t PC_t \tag{11}$$

Tax function

$$T_t = r_t(X_t + M_t) \tag{12}$$

Private medical spending function

$$MC_t = mc_t GNI_t \tag{13}$$

Definition of money gross national income

$$GNI_t = GDP_t + NFI_t \tag{14}$$

Definition of government saving

$$SG_t = T_t + TPFATG_t - GC_t - TR_t \tag{15}$$

Definition of government consumption

$$GC_t = WF_t + ME_t + ED_t \tag{16}$$

Educational spending function

$$ED_t = ed_t \sum_i f_i[P_{i,t} a_{i,t} + j(P_{i,t} a_{i,t} - P_{i,t-1} a_{i,t-1})] \tag{17}$$

"Welfare" spending function

$$WF_t = wf_t\left[1 + j'\,\frac{PC_t - PC_{t-1}}{PC_t}\right]GNI_t \tag{18}$$

Definition of foreign saving

$$SF_t = M_t + TPTAFI_t + TPTAFG_t - X_t - TPFATI_t$$
$$- TPFATG_t - NFI_t \tag{19}$$

Definition of imports

$$M_t = MK_t + MN_t \tag{20}$$

Function for imports of capital goods

$$MK_t = mk_t(I_t + D_t)\sqrt{R_t} \tag{21}$$

Function of imports of noncapital goods

$$MN_t = mn_t(C_t + GC_t + X_t)\sqrt{R_t} \tag{22}$$

Definition of foreign currency reserves

$$R_t = R_{t-1} - SF_{t-1} + B_{t-1} \tag{23}$$

Definition of capital stock quality

$$qk_t = \frac{K_{t-1}^*qk_{t-1} + IMN_{t-1}^*qm + IPN_{t-1}^*qp + IHS_{t-1}^*qh - D_{t-1}^*qd}{K_t^*} \tag{24}$$

Composition of real net investment

$$I_t^* = IMN_t^* + IPN_t^* + IHS_t^* - D_t^* \tag{25}$$

Definition of real imports of capital goods for nonhousing purposes

$$IMN_t^* = MK_t^* - MH_t^* \tag{26}$$

Function for real imports of capital goods for housing purposes

$$MH_t^* = mhIHS_t^*(1 - sh_t)\left(\frac{MK_t}{MK_{t-1}}\right) \tag{27}$$

Housing construction function

$$IHS_t^* = hs_t(P_{\text{males }20-54,t} - P_{\text{males }20-54,t-1}) \tag{28}$$

Definition of money gross domestic product

$$GDP_t = C_t + I_t + D_t + GC_t + X_t - M_t \tag{29}$$

Definitions of real gross domestic product

$$GDP_t^* = C_t^* + I_t^* + D_t^* + GC_t^* + X_t^* - M_t^* \tag{30}$$

$$GDP_t^* = \frac{GDP_t}{pp_t} \tag{31}$$

Definition of real private consumption

$$C_t^* = \frac{C_t}{pc_t} \tag{32}$$

Definition of real net investment

$$I_t^* = \frac{I_t}{pk_t} \tag{33}$$

Definition of real government consumption

$$GC_t^* = \frac{GC_t}{pg_t} \tag{34}$$

Definition of real imports

$$M_t^* = \frac{M_t}{pm_t} \tag{35}$$

Definition of real imports of capital goods

$$MK_t^* = \frac{MK_t}{pk_t} \tag{36}$$

Terms Whose Values Are Changed by Eradication

C	private consumption expenditures
ED	education expenditures
es	elasticity of output with respect to skilled labor
eu	elasticity of output with respect to unskilled labor
GC	public consumption expenditures
GDP	gross domestic product
GNI	gross national income
I	net investment (private and public)
IHS	housing construction expenditures
IMN	imports of capital goods for nonhousing purposes
IPN	expenditures on locally produced capital goods for nonhousing purposes
K	capital stock as of beginning of year
LS	number of skilled workers
LU	number of unskilled workers
M	imports
MC	private expenditures on medical care
mc	propensity to spend on private medical care
ME	expenditures on malaria eradication campaign
MH	imports of capital goods for housing purposes

MK	imports of capital goods
MN	imports of noncapital goods
ms_i	morbidity rate among skilled persons in the ith age-sex group
mu_i	morbidity rate among unskilled persons in the ith age-sex group
P	population
PC	population of equivalent consumers
PDI	private disposable income
pp	implicit deflator for gross domestic product
qa	index of allocative quality
qk	index of quality of capital stock
qs	index of quality of skilled labor
qu	index of quality of unskilled labor
R	reserves of foreign currency as of beginning of year
SF	foreign saving
SG	government saving
sk_i	proportion in the ith age-sex group having received secondary education
SP	private saving
T	tax receipts
TR	transfer payments from government to individuals
WF	"welfare" expenditures (public consumption expenditures excluding those for malaria eradication or education)

Terms Whose Values Are Independent of Eradication

A	constant term in production function
a_i	rate of school attendance in the ith age-sex group
as_i	rate of school attendance among skilled persons in the ith age-sex group
au_i	rate of school attendance among unskilled persons in the ith age-sex group
B	sum of balancing items in international accounts
D	depreciation
ed	education expenditures per standard pupil
ek	elasticity of output with respect to capital
f_i	ratio between per pupil expenditures in the ith age-sex group and expenditures per standard pupil
g	government transfer payments per equivalent consumer
hs	housing construction expenditures per additional male aged 20–54
j	ratio between education expenditures per additional pupil when student population increasing and expenditures per pupil when student population unchanged
j'	ratio between "welfare" expenditures per additional equivalent consumer when population of equivalent consumers increasing and expenditures per equivalent consumer when population of equivalent consumers unchanged
mh	import content of nonsubsistence housing construction when imports of capital goods unchanged from previous year

mk	propensity to import capital goods
mn	propensity to import noncapital goods
NFI	net foreign income from abroad
os_i	rate of nonparticipation as labor input for reasons other than school attendance and morbidity, among skilled persons in the ith age-sex group
ou_i	rate of nonparticipation as labor input for reasons other than school attendance and morbidity, among unskilled persons in the ith age-sex group
pc	price index for private consumption
pg	price index for public consumption
pk	price index for investment
pm	price index for imports
qd	quality of depreciation relative to quality of beginning-of-period capital stock
qh	quality of housing construction relative to quality of beginning-of-period capital stock
qm	quality of imported capital goods (for nonhousing purposes) relative to quality of beginning-of-period capital stock
qp	quality of locally produced capital goods (for nonhousing purposes) relative to quality of beginning-of-period capital stock
r	tax rate
sh	subsistence housing construction as a proportion of total housing construction
sv	marginal propensity to save out of private disposable income per equivalent consumer
$TPFATG$	transfer payments from abroad to government
$TPFATI$	transfer payments from abroad to individuals
$TPTAFG$	transfer payments to abroad from government
$TPTAFI$	transfer payments to abroad from individuals
TTC	terms-of-trade contribution to real national income
V	error term in production function
wf	propensity to spend on "welfare"
X	exports

DISCUSSION

GEORGE H. BORTS: The papers presented by Professors Rottenberg and Barlow point up a paradox which applies not only to medical research but to other types of public investment: it is easier to analyze and identify the economic effects of public action than it is to specify rules for the allocation of resources in the public sector. Professor Rottenberg discusses rules which might be used to determine the optimal allocation of biomedical research resources. He asks whether it is possible to fashion a set of rules to determine the optimum size of the industry and its optimum direction. He has only begun to peel away the layers of the problem.

As Rottenberg suggests, it is easy to fashion incorrect rules and difficult to fashion correct ones. Let me illustrate. Consider a society which own all of its human and nonhuman resources and attempts to maximize the value of output therefrom. Also assume that society employs a single preference function in order to determine marginal rates of substitution among consumer goods both now and in the future. While one could examine the entire medical research industry under the above assumptions and prescribe changes, the examination may not be worthwhile. There are two reasons for this. First, human beings may not in theory or in fact treat themselves as machines. Second, individuals have different tastes, and there is no single best preference function to use in evaluating different final products. The ultimate function to use depends on the choice of the best personal distribution of income. I shall comment on each of these points briefly.

There are many types of medical research which are valuable in our society which would be valueless in the human machine type of society. For example, much of the research on geriatrics would have little value when humans are considered as machines. For this reason I suspect that the human machine norm must be eliminated and alternative methods of evaluation explored. We might regard individuals as maximizing some utility concept related to their families' consumption level. We would then look at the cost of disease in terms of the resources used up. This includes not only potential and actual productivity losses to the diseased and his family, but the resources used in treatment. Each disease might then be evaluated in terms of the costs which it imposes on society, and a reward function derived showing the relative savings from eradication. One implication of this approach is that lingering disease imposes higher costs than suddenly fatal disease, because treatment costs are in part a function of the duration of disease. It may also be possible to show that diseases of middle age are more costly than diseases of childhood.

There is, nevertheless, some question whether such an evaluation scheme is useful. A complication in all evaluation of public expenditure is that the returns from the service are valued differently by different individuals. The amount which they are willing to spend on the service is a function of their income. If the distribution of income among individuals is not optimal, then the reward evaluation system based on this distribution will not be optimal

either. This is a difficulty in all attempts to evaluate public investment projects. Either we must know the best distribution of income or evaluate projects under all possible distributions. I nevertheless feel that the method of procedure suggested is reasonable. If at a later date, for example, the proportion of income going to older people were to increase, then under the suggested method we would increase the social importance of diseases of old age. The economist must use the existing distribution of income as a first step in constructing his scale, and then change it later if the distribution of income should change as a consequence of social policy or economic events.

With these considerations in mind, it is useful to turn to an actual attempt at evaluating the rewards from eradicating a specific disease. Professor Barlow's paper on the elimination of malaria in Ceylon is in many ways a model of the type of investigation which should be carried out for any type of economy. Nevertheless, it consciously adopts the narrow point of view I mentioned earlier; namely, treating individuals as producing machines and focusing on the flow of gross national product which results over time from disease eradication. While the model allows for a reduction in private medical care expenditures, such savings are evaluated, not as an increased flow of current satisfactions to consumers, but as an increase in the contribution to capital formation and future output. The distinction between these flows is ignored, because the model assumes a constant household savings ratio. In fact, the two flows are equivalent only under specific conditions of intertemporal equilibrium in production and consumption which are ignored.

Professor Barlow shows that the main effects of eradication of malaria are an increase in labor force participation and quality, which directly raise produced income subject to the diminishing returns possible in his aggregative production function. There is in addition an upward shift of the production function reflecting an improvement in allocative efficiency. This shift occurs because eradication increases the supply of land available for production. In view of the assumption that land is not a specific argument of the production function, the function must shift upward.

While the initial impact of eradication is a jump in income per capita, it is followed in later time periods by factors which tend to slow the growth of income. These factors may be summarized as follows:

1. The quality of the capital stock is reduced because there is a greater proportion of investment taking the form of housing construction. Housing construction is regarded as a lower productivity form of investment. The explanation for this treatment is not clearly provided. It may be that gross national product is so narrowly defined as to exclude some of the services rendered by residential housing. If so, there is an error in the measure of social product. Alternatively, the restrictions on importation of capital goods suggest that they are more productive than domestically produced capital goods, which include housing. Such a suggestion is illusory, however. Prices can adjust to equalize the income flows from a dollar spent on any type of capital good.

2. Over time the amount of saving which is financed through government surplus is reduced because of increases in government expenditure on education. The increase is caused by the extra population produced by the eradica-

tion program. As a consequence of the reduced level of government saving, the private investment which should have been financed by such saving must presumably fall.

I think that the restrictions of Barlow's model result in an understatement of the economic benefits of eradication. The major restrictions are found under the sources of savings contributed by government and by the foreign sector.

Government saving consists of the surplus of tax receipts and foreign gifts over government spending and transfer payments. Over time this source of saving must be expected to decline in its ratio to the population. The reason is that tax receipts are related to the sum of exports and imports, while government spending is related directly to population size. Further, exports are exogenous to the model and apparently independent of the level of population. A very strange assumption. Imports are constrained by the value of exports and foreign exchange reserves. The size of foreign exchange reserves is endogenous to the model and presumably constrained to be positive. The result is an upper limit to imports such that imports, government tax receipts, and the contribution of government to total savings decline as a percent of the population.

For the same reasons, the contribution to savings from the foreign sector must also decline as a percent of population. The net result of these two factors is that over time there is a continual fall in the share of income saved. The consequence is a continual decline of output per capita. Why, then, does the simulated income per capita appear to rise in the absence of eradication? The answer must be the effects of education expenditures on the productivity of labor. Education is the mechanism by which unskilled labor is converted into skilled labor. Skilled labor has a markedly higher productivity than unskilled, and over time, income per worker and per capita will rise as the proportions of the labor force shift toward the skilled group.

As a consequence of the above restrictions on savings behavior the benefits of malaria eradication are understated. First, I find it very hard to believe that eradication will have no effect on the supply of exports. Second, I find it hard to believe that eradication would leave foreign borrowing unaffected. It may have been realistic in yesterday's Ceylon to rule out foreign private borrowing; it may not be realistic today. As a result, a very important source of savings is denied the economy. Finally, the evaluation of income flows is based on the assumption of unchanged household savings behavior. With the increases in capital productivity resulting from eradication this assumption also leads to an underestimate of benefits.

HERBERT E. KLARMAN: Dr. Rottenberg has chosen a difficult area, one in which economists have not yet done any empirical work. He has written a "think" piece that reflects the application of his background in economics to the medical research literature.

His thesis is simple. The optimum size of the medical research industry is determined by the equality of marginal cost and marginal revenue. In the allocation of research resources among diseases, choices are being made, explicitly

or otherwise. It is better to make the choices explicit and rational rules exist for so doing. He proceeds to state some.

I agree with the thesis as broadly stated and also with much of the underlying argument. I take issue with some of the details of the argument. Dr. Rottenberg and I also seem to differ in our understanding of how medical research resources are actually allocated.

The extent of agreement is substantial. I agree that the value of research depends for the most part on its contribution to health services. I agree that the presence of multiple diseases and competing causes of death in sequence causes difficulties but see no obvious solution. We all recognize the difference in society's appraisal of the value of human life when one known person is involved and when a statistical composite with similar attributes is described.

Several of my comments fall in the category of "Yes, but."

True, expenditures for research should relate to value. But value is determined by future returns, not even partly by past investment.

True, rational choices are superior choices. But that does not mean that research is like ordinary business. Human beings are not discarded when their replacements are able to produce at lower cost, as machines are. In business losses lead to bankruptcy; in research negative findings are useful and a research policy that results in a high yield of successes is likely to be too conservative.

True, intangible benefits of health must be recognized in allocating research resources among diseases. However, the treatment of intangible benefits to which I object is not uniform valuation per case but the assumption of a constant ratio between intangible and tangible benefits for all diseases.

True, research is not a random process. But serendipity ensues when an investigator is prepared to see the unexpected and to exploit it.

I disagree with Mr. Rottenberg in certain areas in which neither of us is qualified to say the last word. For example, I do not believe that physicians are casual in administering anticoagulents. Rather, physicians disagree on the efficacy of these drugs in treating heart disease. Moreover, some physicians are conservative in prescribing drugs, emphasizing the avoidance of harm.

Nor do I believe that the treatment of paresis poses a dilemma today. In fact, penicillin is a specific remedy for syphilis and for preventing its consequences.

Perhaps more important is my belief that it is not useful to pose prevention, diagnosis, treatment, and rehabilitation as alternatives, preferred in descending order. It is always necessary to know the specifics of a program, its costs and probable returns, with the opportunity for success in research constituting a large element of the returns. Prevention is preferable to cure if the costs are the same, not necessarily otherwise.

I now turn to how medical research resources are allocated under the study section system of the National Institutes of Health, as I know it. In reviewing grant applications, study sections make no explicit allowance for the opportunity that exists to discover a missing piece of knowledge. The primary emphasis in reviewing an application is on the quality of the design of the research product. Relatively less attention is paid to the capabilities of the senior investigator, who may be unknown.

The amount of money allocated to a project is determined, not by its value, however measured, but by how much is required to carry it out, given its approval for scientific merit.

All of us know that certain areas of medical research may be overfinanced. Drug trials for cancer are a good example, demonstrating our acquiesence in the principle of seniority.

Prostheses for amputees are a good example of the underfinancing of research. Were it not for the work of the Veterans Administration, the state of this art would be deplorable.

It seems reasonable to suppose that initiative on the part of some agency with an objective view of all of medical research would serve to reduce such misallocation of research resources. One current difficulty lies in the lack of a suitable mechanism. Research grants reflect the wishes of applicants. Contracts are intended to fill gaps, to meet the needs of the field. Under a contract the government not only exercises initiative in the selection of the problem to be investigated but its representatives closely supervise the conduct of the research. By law the contract is considered to be an extension of the arm of the government agency that pays for it, giving the outside investigator no leeway.

Another difficulty lies in certain types of uncertainty to which medical research is subject. I do not mean to dispute here the contention that medical research is best conducted by disease category because investigators are specialized. Obviously, it is necessary to have some classification system in order to administer research. For a given class, however, diminishing returns apply, unless an opportunity for breakthrough beckons and the productivity curve is lifted. It is thus possible to establish some priorities. The uncertainty referred to has to do with the fact that research carried out in one field is sometimes applied in another. Chance is one factor. But design is also involved. For example, much of the current research on transplanting the kidney from a cadaver is a necessary preliminary to transplanting the liver and the heart, for which live donors cannot be used.

Rottenberg's suggestion that the problem of crossover discoveries be handled by upward adjustment of their value in the field in which the discovery is made seems to be more readily applicable to crossover by design than to crossover by accident. I can see no way of handling the latter problem, short of studying what has happened in the past. Empirical research of this type is badly needed.

Finally, I have a question to add. What constitutes "payoff" in medical research? Suppose a discovery is made but is not acceptable and is not applied. How much weight should be assigned to the probability of application?`

The paper by Drs. Long and Feldstein is noteworthy for attempting to minimize the total costs of caring for a given number of obstetrical patients in a community: the costs of hospital care, travel costs, and the costs of inconvenience. For the first time the three factors are brought together and their effects quantified. With economies of scale posited for the obstetrical unit, travel costs which increase with the size of unit serve as partial offsets. Inconvenience costs turn out to decrease with the size of facility and can serve as compensation for the reduction achieved in the total number of beds.

The key conclusion of the paper is that higher occupancy rates can be

achieved for a hospital system, without physically combining units, by allow-ing more patient transfers. The reason is that variation in the demand for ser-vice—and the relative amount of stand-by capacity—is a function of the size of the population served, not of the size of facility. Small facilities that do not serve designated subgroups of a population are equivalent to a large facility that serves the entire population.

First, I have some comments on each of the three factors in the model.

Hospital Costs. Throughout the paper, except in the preparation of the final estimates, economies of scale are posited, owing to the presence of factor indi-visibilities and the tendency of the rate of occupancy to increase with size. The latter result is obtained from two mathematical distributions employed in operations research that fit variations in obstetrical patient load—one for ad-missions and one for duration of stay. I strongly suspect that short duration of stay is itself a factor in a low rate of occupancy, owing to the enhanced difficulty of avoiding lags in filling beds.

Travel Costs. The allowance for travel is too low, since time is costed but not transportation expense. The doctor's extra travel time to another hospital during working hours is neglected. The curve as drawn may be too steep, if the first part of any trip is in fact disproportionately costly.

Inconvenience Costs. It is not clear what these represent. First, they are said to "include an assessment for reduced quality." Later inconvenience is said to occur when the absence of facilities at the chosen hospital necessitates a departure from regular practice. In the first case, $500 may be too low a value, particularly if it reflects the extra risk that the doctor may miss the de-livery. In the second case, $250 may be too high if the doctor has a staff ap-pointment at the second hospital and the patient's medical record follows him. Perhaps a further specification of inconvenience and the purposes of compen-sation is called for.

It occurs to me, too, that the number of persons who require special han-dling may be overstated. It is safe to introduce some adjustment in the dura-tion of patient stay, once the importance of the religious factor is recognized. In a sense hospital bed capacity is slightly understated by the exclusion of de-livery beds from the bed count while their patients are included in the hos-pital's census.

More serious is the treatment of obstetrical facilities as if they were sepa-rate entities. The fact is that we have abandoned maternity hospitals and in-corporated obstetrical units into general hospitals. If is important to under-stand the reasons for this development. Patients require consultation from medicine, radiology, laboratory, and pediatrics. Being a complete hospital strengthens the hospital both as a service and an educational institution.

The treatment of obstetrical facilities as separate entities may prove to be misleading for planning. The basis for planning, as given in this paper, is non-profit ownership and motivation. Perhaps no other basis is visible where third-party payments are relatively small. For medicine and surgery, however, many of us have come to believe that under conditions of third-party payment use is influenced by the supply of beds. If so, a divergence exists between the interests of the community and the individual hospital, and intervention through plan-ning is indicated.

The separate treatment of obstetrics leads one to think that hospital planning should be done in terms of particular services. That is necessary where professors or chiefs of departments manage their own separate services. Where beds are used by many categories of patients alternately, this is not necessary. As the authors recognize, the segregated use of facilities should be avoided because it is costly.

Most serious to my mind is the failure of Long and Feldstein to explore more fully the implications of a patient transfer policy. The authors recognize that physicians' staff appointments are involved. My point is that while multiple-staff appointments for physicians have a proper place in coordinating the use of hospital facilities, they also have certain limitations. In general, the organization of the medical staff by a hospital and control over the staff list and the privileges of members serve as guarantees of the quality of care. Secondary, or courtesy, appointments make no contribution to the quality of care in a hospital or to its development in other respects. It might be noted by contrast that in Europe hospitals have full-time medical staffs, closed to doctors in the community. This system is said to afford advantages in quality of care and in reduced use of services. Finally, it has been suggested (by Robert Sigmond) that a physician on the staff of several hospitals will attempt to play off one against the other and insist that he must have access to a bed when he wants it and where he wants it.

A final comment on planning may be in order. Hospitals serve local areas. It is notoriously embarrassing to check one's forecasts of population for small areas. Dr. Feldstein has shown elsewhere that the best predictor of future per capita use is still past use. I conclude that the importance of accurate forecasts of demand for buildings with a life of forty years or more is overrated. Planning agencies would do well to devote greater efforts to developing devices for flexibility.

PETER NEWMAN: Robin Barlow's well-written paper on measuring the economic effects of malaria eradication is, as far as I know, the first example we have of a comprehensive attack on this difficult problem. It is important to make clear just what are the advances that this work contains, for its innovations are to some extent concealed by its modesty. In the beginning, before economists came on the scene, there was a lovely and fragile innocence in the attempts to evaluate economic aspects of eradication and near-eradication. *Post hoc ergo propter hoc* was the order of the day and the useful distinction between gross and net increments to output was honored mainly in the breach. Later, especially as economists and others began to mutter from the sidelines about the adverse consequences of sudden acceleration in rates of population growth, these estimates began to improve but even so still failed, in Barlow's words, "to take into account the multiplicity of economic effects" of eradication.

Neither was the economist's contribution much better. Quite apart from the tendency (common to much of their work in the economics of health) to underestimate the constraints enjoined by technical and organizational realities, many economists (including myself) have been altogether too glib in attributing any and all spurts in population growth in the tropics to the effects

of malaria eradication; moreover, they have tended—at least in practice if not in theory—to regard all such accelerations as necessarily inimical to economic growth, almost regardless of the individual country concerned. So it is not surprising that public health workers, while rightly uncomfortable about the adequacy of their own attempts at evaluation, are disposed to regard the economists' views with suspicion, as not containing the whole truth of the matter. The work of mine referred to by Barlow and his own research may be looked on as systematic attempts to help fill these two glaring gaps in our knowledge of the demographic and economic consequences of eradication, respectively.

My study proposed a method by which we can split the causes of population growth into two additively separable components; namely, those due to the eradication of malaria and those due to all other causes taken together. It would be idle to pretend either that the method cannot be improved upon substantially, or that it can be applied for every country in which malaria is or was a serious health problem (considerations of data availability alone preclude that); but there is some independent evidence that the model is reasonably appropriate for the problems at hand, at least for some countries.

The leading results produced by application of the method to Ceylon and to Guyana were that in Ceylon, malaria eradication accounted for 60 percent of the acceleration in population growth that took place from the prewar period (actually 1931-46) to the postwar period (1947-60), while in Guyana the corresponding—and I believe much less reliable—estimate was 40 percent. Expressed in different terms, these results imply that in each country 0.7 of a percentage point of the postwar crude rate of natural increase (CRNI) of population was due wholly to malaria eradication; thus, of the remarkably constant CRNI of about 2.8 percent in Ceylon from 1947 to 1960, one-quarter can be attributed to eradication.

Faced with such magnitudes as these, the investigator of the economic effects of eradication cannot apply the by now standard techniques devised by Dublin and Lotka, as refined by such workers as Fein and Weisbrod. These methods assume that, while the economy certainly affects the newly saved survivors of the disease, the numbers of those saved are relatively so few that the economy itself is not significantly affected. This is the classic assumption of partial equilibrium analysis, and it implies, for example, that the real and money wage rates can be taken as data, independent of the effects of eradicating the disease. It is a reasonable assumption for the effects of reducing such diseases as tuberculosis and poliomyelitis in modern rich nations, but quite unreasonable for such a disease as malaria in such a country as Ceylon, where eradication raised the population growth rate by one-third. Here it is essential that the investigator devise an adequately general equilibrium model, that will permit the causal effects to go both ways, and not only one way as with the Dublin-Lotka methods. It is the chief merit of Barlow's pioneering work that it is the first to recognize fully and to deal adequately with this general equilibrium problem—a problem several orders of magnitude more difficult than the inadequate partial approach.

Turning to the model itself, obviously the difficulties of data collection and parameter estimation have forced Barlow to make some heroic simplifications,

whose precise nature and implications are hard to judge on the basis of the necessarily abbreviated version in this paper. Although there seems little doubt that the model captures some of the basic themes of interaction, there is certainly room for discussion about variables that are omitted (such as wage rates) and about the "malarial" status of others that are included. In particular, while sympathizing with the author's desire to make exports exogenous of malaria, I do not find his arguments fully convincing. While it is true that the formerly endemic malarial zone produced and produces few export crops, the formerly epidemic zone produces large quantities of coconut and rubber, and another severe epidemic such as that of 1935 might have noticeably affected exports. In a country whose main dynamic was and is the behavior of exports, the assumption that malaria eradication brings no benefit to exports seriously reduces its power to help the economy.

When Barlow comes to run some more simulations with alternative assumptions, he might also investigate the consequences of using my upper and lower estimates of the effect of eradication on population growth rather than the point estimates alone. Aside from this, I have some reservations about the procedure adopted in his simulations. As I understand it, several of the parameters of the model are adjusted for each year of the "sample" period in order to reproduce exactly the actual values of per capita income. The "noneradication" simulation follows the same plan, except for changes in the variables directly affected by eradication. The difficulty in this procedure will come when Barlow pushes forward the simulation to periods in the future. Then the simulation describing the actual flow of events will have to assume parameters for the future based on past behavior, and the comparison between the two simulations will necessarily involve a composite both of the "real" effects of malaria eradication together with the rather extraneous differences due to imperfectly known parameters; the possibility exists of a rather sharp jolt in the comparison series as one moves from present to future.

In my view it would be preferable to estimate average values for the parameters from the sample period (the estimation will often involve crude guesses), and then use this estimated model to generate "predicted" values of per capita income for the whole period from 1947 to 1977. These can then be compared directly with the similar estimates obtained by assuming noneradication, and the resulting differences can throughout be ascribed solely to the effects of eradication and not also to the divergence between model and reality.

I would like to close by first of all congratulating the author for adding to his merit by avoiding the besetting sin of the economist (from which the other two papers in this session are not entirely free); namely, the derivation of "policy conclusions" from a model which is necessarily an early and simplified attempt. Second, I would like to plead with the public health workers interested in this problem of evaluation to recognize that it is deceptively difficult, and to understand and sympathize with the complexities contained in such work as Barlow's. These are not introduced as mere economic pyrotechnics but are inherent in any adequate handling of the problem. Simple-minded models simply will not work.

ECONOMIC ANALYSIS OF WATER RESOURCE PROBLEMS

NONMARKET VALUES AND EFFICIENCY OF PUBLIC INVESTMENTS IN WATER RESOURCES*

By A. ALLAN SCHMID
Michigan State University

The task is to form questions directed to predicting the performance of the economy with reference to the production and distribution of certain water derived goods which are often alleged to be inappropriate for market exchange. These include reduction of threats to human life, health, scenic beauty, and preservation of historic sites. For convenience I shall often lump them together and call them environmental goods.

It is difficult to say anything new. Still, a great deal of confusion surrounds discussion. Perhaps if we can sort out our concepts and what we know systematically, it would be worthwhile.

What are the characteristics of these goods? One variation of the nonmarket concept is the label of intangible goods which is used in practice by the federal resources agencies and is often defined in the literature as a good that "can't be priced" or cannot be reduced to a common denominator (incommensurable).[1] How can a price be put on a human life or a wilderness stream experience? The question might better be, how can one avoid doing so. The provision of most goods involves an alternative use foregone. However the decision is made, a choice of resource use involves an opportunity cost. Any law or standard for flood prevention or drinking water involves an exchange of that commodity for another that might have been obtained.[2] There is an implicit exchange ratio or price for every act for which an alternative exists. Any administrative standard has the effect of pricing the good obtained in a common denominator with the goods foregone. If two goods (A and B) are exchanged, they can each be expressed in terms of the other, and if a third is exchanged for either of them, then both (A and B) can be expressed in terms of the third commodity.

[1] For example see, Roland N. McKean, *Efficiency in Government Through Systems Analysis* (Wiley, 1958), p. 58.

[2] Harold A. Thomas, Jr., makes the same point and expresses it as a benefit-cost ratio, "The Animal Farm: A Mathematical Model for the Discussion of Social Standards for the Control of the Environment," *Q.J.E.*, Feb., 1963, pp. 143-48.

Prices are implicit in the process of choice, and the relevant question is the rules surrounding price generation.

In popular usage intangibility is sometimes used to mean that a good has nearly infinite value. A good is walled off from its alternatives and no further analysis is needed, since no matter what is foregone the commodity is assumed to have a higher price. The possibility of such evaluation cannot be theoretically denied, but one can observe that there seem to be limits in past choices, and that some risk to life, etc. is tolerated in order to gain other useful goods.

It is possible that people may evaluate a good so highly that they isolate it and will not exchange it, rejecting all bids, or, if they do not own it, will devote all of their resources to achieve it. But a market is perfectly consistent with the above. If people want to withhold a commodity from exchange, this is different than saying the good cannot by its nature be marketed. This does not deny a common denominator but only that the price is high. Societies may wall up certain commodities and prohibit their exchange. The exchange foregone, however, is its price.

In practice, people often make a simultaneous group pricing of some good and say that all other bids are thereby rejected. It appears as a prohibition of exchange, but the very act is to trade the good obtained for the alternative foregone and thus there is an exchange ratio even if no goods change hands or ownership.[3] This pricing can be observed even though the total value of the good cannot. This is the concept of a reservation price, only in this case it is simultaneous and jointly arrived at.

Another variation of the intangible concept is the idea that some goods cannot be measured. Tied up with this is a new distinction between quantity and quality. The claim is that we have been obsessed with the quantity of goods like steel and now should consider the quality of the environment; e.g., flood threat or scenic views. I find this distinction confusing. One must have a taxonomy before one can count things. Quality in the sense of product definition and categories are essential to quantification. You cannot have one without the other. The above claim should be reworded to say that we should demand more units of goods like views and fewer units of goods like steel. The label of intangible in this connection tells us more of our ignorance than the character of the good itself. It usually means that we have not thought enough about the utility of a good view and therefore do not know if we have more of them or not or whether we would be willing to exchange anything to get more.

[3] Karl Polanyi, *Trade and Market in the Early Empires* (Free Press, 1957), Chap. XII and especially p. 266.

We may be vaguely aware of a want and not be sure how to satisfy it (what products and services will fulfill it). We may feel that certain scenic views are conducive to mental health. Yet when we do not have the information, we just claim that the product is intangible when we really mean that we are ignorant of its connection to desired character-istics of man. When knowledge becomes available, the product ceases to be intangible, and views can be considered along with tons of steel. It is true that we lack knowledge of the role of various environmental goods to our personality. It is just as true of cars as it is of views. We do not know much about how cars affect us; yet we do not regard a car as a mystical intangible.

It is confusing to talk of a change in the characteristics of water commodities as changes in product quality. They are simply different products. It might be more useful just to consider water with different chemical and biological content as different products, like steel, stain-less steel, and aluminum are different products and not changes in the quality of metals. We can then speak of choice of product mix by con-sumers and firms in the usual way. Each water product then has a different cost and demand schedule and the interactions such as cross-elasticities can be analyzed.

I have been talking about price and not value. Economists know the price of many things and the total value of nothing. This is just as true of steel as of views. People might pay more than they now have to for a shoreline campsite at a national park, but there probably is also con-sumer surplus in steel. I do not know how to determine which is the greater, and if I read I. M. D. Little correctly, the question is not only unanswerable but also meaningless (for several reasons including the fact that people do not know themselves).[4] An administrative standard (e.g., water quality) set by government does not establish the total value of any commodity any more than a market transaction does. Both procedures may be described in terms of price, however.

To summarize briefly, I have argued that the concepts of intangible and incommensurable products are devoid of usable analytic content. There is, however, another aspect of goods that seems useful for pre-dicting market performance. This is public goods.

Public Goods

There are two components of the concept of public goods discussed in the literature:[5] (1) $MC = 0$ for some group; i.e., consumption by one individual does not reduce utility to another; (2) potential users

[4] I. M. D. Little, *A Critique of Welfare Economics* (Oxford Univ. Press, 1957), Chap. X.
[5] See Paul Samuelson, William Baumol, and Mancur Olson.

cannot be excluded if the product exists at all. These two need not go together.

If a certain water analysis is maintained in a reach of a stream so that those who live along it are not offended by smells and floating objects, the cost of obtaining this product is not affected by the number of consumers. If this product must be produced by treatment of an upstream waste discharge, the marginal cost of another beneficiary is zero. Here product definition and units must be kept straight. The product is a flow of water of X analysis and the units are numbers of users viewing and smelling. If one person pays enough to cause the water to be retained at X quality, it is available to all using it for aesthetic purposes at no extra cost, and also in this case they cannot be excluded if the product is produced at all. This creates problems for bilateral market bargaining when the benefited group is so large that an individual does not affect the outcome. The individual knows that if others provide the public good, he will benefit whether he pays or not. The individual has no market incentive to reveal his preferences and avoid being a free rider, and incidentally no reason to misrepresent his preferences or make strategic moves. Bilateral market bargaining alone could then mean that the product will not be produced, although it can be imagined that there is a group of buyers who have positive marginal utility for the good and would be willing to share the cost. This does not mean that if the group can effectuate its demand through an agent (by some political process) the good then subsequently cannot be exchanged by markets.

It should be kept in mind that a right to dispose of a given range of water characteristics (e.g., the right to lower oxygen content from 10 to 0 p.p.m.) can produce a public good for fishermen if kept at 10 p.p.m. or a public good for polluters if lowered to 0 p.p.m. If we allow group bids from fishermen, can we disallow group bids from polluters? I can envisage two government agencies, one representing fishermen and another industry, bargaining in a market (perhaps an auction) for the right to water. While each agency relates its constituents to each other in a political process, the two groups could be related to each other in a market. The power to buy as a group is a valuable property right, the distribution of which will affect demand curves just as surely as income distribution. Can a society afford to indiscriminately give the right to buy as a group?

There are implications of public goods for appropriation laws which are still commonly used to distribute ownership to various natural resources. Many environmental products have rights established by the first user. Many scenic views have been interpreted by the courts as belonging to the first owner of land who overtly acts to use them;

e.g., the first owner to cut off a view or irrigate becomes the owner of the right to dispose of it as he wishes. Those who had been using the resource for views, etc., now find that they had been using the property at the suffrance of the now established owner. While economists have generally taken a dim view of appropriation procedures, if they are retained at all, perhaps groups should be allowed to file for rights to the resource in its existing character and not require an overt act and taking for the ownership to vest.

Externalities

There is another concept that must be examined along with public goods. This is "externality," which is defined in the literature as an interdependence of production or consumption functions which is not adequately taken into account (compensated for) by decision-makers. It is often said that if a firm pollutes a stream without accounting for the damage to other users, this is an externality. There is a strong tendency to conclude that something should be done about it in the name of social welfare. I shall argue that the concept of externality does not describe the character of the product. Rather it describes the existence of Pareto-better trading possibilities. Since such trading possibilities depend on a prior definition of property rights, the existence of an externality is no guide to the choice of property rights.

The argument turns on what is meant by inadequate accounting, compensation, or damage. The existence of externalities is not meant to refer directly to the question of income distribution. If fishermen have been using property at the suffrance of its owner, they cannot complain when the ownership rights are later exercised. Such is the plight of the poor man—he just does not count because he owns very little that is a cost to other people.

Those that draw implications of externalities for property rights and economic efficiency and welfare have something quite different in mind. The argument is that because of externalities the products of pollution, e.g., paper or chemicals, are too cheap, water is misallocated and national income reduced. Buchanan and Stubblebine have furnished a clear definition of a relevant externality; i.e., a Pareto-relevant one. "An externality is . . . Pareto-relevant when the extent of the activity may be modified in such a way that the externally affected party, A, can be made better off without the acting party, B, being made worse off."[6] The concern is not with the poor man, but those with money whose demand is not communicated for some reason. If the owner of the water (e.g., a paper firm) does not get the message that

[6] James Buchanan and W. Craig Stubblebine, "Externality," *Economica*, Nov., 1962, p. 374.

some fishermen would pay enough to reduce pollution, a chance is missed to make both the paper firm and the fishermen better off through trade. This is old hat and the main basis for policy recommendations to remove barriers to trade for any kind of product. Because fishermen may be willing and able to buy water but cannot do so through bilateral bargaining, it follows that group bargaining must be allowed, indeed encouraged, if externalities are to be removed and welfare maximized. Public goods require some alternative to bilateral bargaining if demand for them is to be communicated, all possibilities for Pareto-better trades exhausted, and externalities removed.

A distinction is made in the literature between technological and pecuniary externalities. I shall not repeat this well-known argument but merely note its conclusion that something should be done about the former while the latter should be ignored. The example referred to above is a case of technological externality. The paper firm affects the fishermen's production function directly, and if they can trade, there may be Pareto-better possibilities (if group decision-making costs are not too high). But what about other polluters who were not noticed before or the buyers of paper products who will find higher prices—a pecuniary externality. Of course, if paper buyers have no rights in existing prices, then by definition the allowance of fishermen group buying is Pareto-better and the whole relevant world is better off.

In historical practice, policy has had a split stance on this matter. In general, firms are prohibited from avoiding the effects of price changes by buying off competitors. Yet, a limit is acknowledged to the extent to which major groups can get clobbered in price changes. This is evident by things like agricultural price supports which in effect acknowledge a farmer's property right in prices. If things move too fast, they destroy property without due process. We grant rights in market access and good will even though interference with them does not affect the owners' production function.

There is no theoretical efficiency, pure competition, or welfare guide to these questions. Efficiency depends on the relevant input-output categories, and it is collective judgment expressed in property rights which determines relevancy. If property rights are accepted, then Pareto-relevant externalities are defined, and it is logically consistent that they must be removed whether technological or pecuniary.

A man is affected whether someone else pollutes his water (technological) or bids up the price of an input or consumption good (pecuniary). These are surely different things, but their welfare implication depends on a judgment of who is to count in a society and what performance of the economy is wanted and the factual relationship of alternative property rights distribution to this performance.

One cannot talk of Pareto-better trades unless property rights are first established to say who counts. Depending on the property rights, there are all kinds of Pareto-better trades possible which would produce a whole range of price sets (i.e., relative prices of paper, chemicals, fishing, and views) and other performance results in terms of growth in per capita income, inflation, employment, kinds and variety of available products, and mental health.

The above is well known; yet the popular literature is replete with confusion and some of it creeps into professional writing. It is common to hear that we must weight efficiency values and other noneconomic values. While one cannot talk about efficiency unless the input-output categories have been determined by property rights to specify just what and whose values shall count, the confusion persists.

There are implications of this for redirection of analytic effort. We need empirical measures, not just of possible advantages of trade with given property rights, but measures of interdependence in the sense of how an individual is affected in various lengths of run by changes in property rights. For example, what happens to an individual if he gives up a current property right or does not attempt to gain the initial vesting of a right to a newly valuable resource? Can he hope for a future increase in productivity of his remaining property thereby? The first step is obviously to measure the extent to which public water investments affect income distribution, and efforts such as Haveman's are to be applauded.[7] We need data on the impact of the Pareto-irrelevancies as they now exist.

It is sometimes asserted that all of these nasty problems of property rights choice can be avoided. The proposition has been advanced that "if the party imposing external diseconomies and the party suffering them are able and willing to negotiate to their mutual advantage, state intervention is unnecessary to secure optimum resource allocation."[8] And further, while income distribution is affected, the allocation of resources remains the same. The key to the proposition is the notion that loss can be measured in money terms as profit differences. When both firms enjoy a noncompetitive profit or a site rent advantage, the above proposition can be demonstrated handily. But what happens if the two parties are both tending toward long-run competitive equilibrium?

Assume A produced an effluent which damages B. Since neither enjoys excess profits or rents, in the long run A could not afford to pay damages to B nor could B afford to pay A to refrain from polluting.

[7] Robert H. Haveman, *Water Resource Investment and the Public Interest* (Vanderbilt Univ. Press, 1965).
[8] Ralph Turvey, "On Divergences between Social Cost and Private Cost," *Economica*, Aug., 1963, p. 309.

The firm who has the property right will remain and ignore external effects and the other firm will go out of business with loss of its fixed assets.

Under Buchanan's definition this effect of a new use of water for waste disposal by firm A would be Pareto-irrelevant, since no trade can make them both better off. Thus, externality theory, which is the logical extension of a given set of rights, is not useful for questions of the original vesture of the property right to a resource that becomes newly useful.

If the two firms are marginal to the industry, then the effects on national income either through changed quantities or prices would be negligible. But what if we are talking about the distribution of property rights between whole industries for a nation and not between two firms on a single river? When the question of national pollution standards is discussed, this is the issue. Then the choice is between two sets of relative prices.

The original vesture of property rights can make a great deal of difference in at least two other instances. One involves the situation where the initial resource ownership bulks large enough in the owner's total income that it affects his choice of products. If a poor man has access to a public lake, he might not be willing to give up that right for less than $1,000. Yet if he does not have the right, he might not be willing to pay anything for it. The initial distribution of the property right will affect the eventual use of the resource even if there is a market and Pareto-relevant externalities are absent.

The second case where the initial distribution makes a difference is where the cost of collecting the payment offsets any superiority in willingness to pay that a group has over another buyer. For example, group A would be willing to pay more for the resource than it is worth to individual B. Yet A has the costs of reaching a group decision and of collecting payment. These costs may be great enough so that when subtracted from the total amount they are willing to pay, they could not meet B's reservation price. Yet if the resource were owned by A, then B could not afford to purchase it. This can be summarized as the familiar fact that income and property rights distribution affects demand functions.

Introduction of New Products and Nonmarginal Change

Let us look specifically at some of the practices now in use to handle the products under consideration. The Ad Hoc Water Resources Council in setting evaluation standards for recreation benefits uses the term intangibles to refer to protection of wild areas, areas of unique beauty, areas of scenic, historical, and scientific interest and preservation of

rare species. While it is not established that these are public goods or have other features relevant to the private or group choice question, the Council recommends that these be described narratively and estimates made of additional project costs or reduction in other project benefits necessary to secure them.[9] This recommended form for presentation of data on public goods to Congress is the old "it must be worth as much" concept.[10] It is frequently lamented that we know the costs but not the values of these goods. I will argue that we do not know either one if the new products are nonmarginal. This is just as true for private as public goods.

What are the relevant costs? It will not do just to add up the extra construction costs or the market value of other project benefits foregone. We need to know something of the value of all relevant alternative uses of the water and not just project uses. The opportunity cost of water is a function of water derived product prices. The existing prices have found their relative position in a field where the new water product (e.g., preservation of a rare water bird) has been absent. If we are concerned with only one project and a few birds, this is not a serious problem. Yet if we are talking of embarking on major new programs to introduce a number of new products, we are going to have to estimate how these will affect the price of paper, chemicals, etc., and therefore the opportunity cost of water for environmental products. The introduction of new goods, whether public or private, has always played havoc with pure competition and welfare theory and the measurement of historical economic growth. The existing prices and opportunity cost may be irrelevant.

Neither the demand side nor the supply side stands still. People are deciding both at once when contemplating major new investments in environmental products. The eye must be kept directly on performance. People are making a choice broadly affecting the availability and relative prices of various groups of products. Prices here are a result of the decision, not a guide and dictator of it. Analysis could present alternative performance levels (of differing product mix) and suggest the property rights necessary to obtain the price sets consistent with this performance results. Let the price set anticipate the performance and not take prices as given and circumscribing.

Another practical implication of the above is with reference to the multiagency review that is now required for federal water investments. Much agency time is spent on review of another agency's projects. Yet there is a need for broad program reviews from the points of view of

---

[9] *Evaluation Standards for Primary Outdoor Recreation Benefits*, Supplement to Senate Document No. 97 (June 4, 1964).

[10] For example, see McKean, *op. cit.*, p. 63.

various agencies communicated to Congress and not just the limited project by project comment.

Another consideration in deciding which costs are relevant is economies of scale. The public may reject a new product if introduced incrementally which might be purchased if scale economies were significant and could be anticipated. There are scale effects on the demand side also. There are important threshold points which need analysis. This may be especially important in safety and health products where it is hard for individuals to perceive such things as a half of one percent decline in an area's death rate.

It does no good to assume that a resource investment program is always marginal with no effect on prices and marginal utility of money and has no scale effects or public goods aspects. We must get busy with empirical analysis of the impact of these investments and let the people decide how they want the effect to fall. No unique maximization of national welfare emerges from such a messy world where tastes neither stand still nor are communicated infallibly. Surely in the expression of demand for a public good a congressman is never sure he was defeated because of his stand on the Central Arizona Project or Vietnam. But then the private producer is never sure from price data alone just what features of his product consumers did not like. The rock of consumer sovereignty that we all cling to is slippery. While environmental products can be expressed in terms of price, this does not mean that any price is a criterion. Price is a director of economic activity, but also a reflection and result of choices made at different levels. Economic analysis cannot help people avoid the weighty questions of who should count and what they should make of themselves, but it can indicate what difference various property rights can make.

Perhaps the new research directions called for here can be made clearer by a bit of historical perspective. Keynes taught that bilateral bargaining could bring equilibrium at less than full employment. Each person takes the field of opportunity as given and maximizes within it, but in sequence the field becomes worse until total performance is satisfactory to few. To get a different performance requires a group action to redistribute and restructure the power of decision and where costs and benefits fall. Investment that was unsound for the individual related by existing market rules was sound if coordinated differently.

Recently Boulding and Singh argued that bilateral bargaining may bring an equilibrium of prices and supplies that does not maximize growth;[11] that is, such bargaining could not achieve the price sets most conducive to growth. To get the desired performance may require

[11] Kenneth E. Boulding and Pritam Singh, "The Role of the Price Structure in Economic Development," *A.E.R.*, May, 1962.

group action. Acting individually and related by a market, people may not be able to achieve the price set consistent with the performance they want. Modification of property rights or direct government administration may be required to produce the desired price set.

The same conception is applicable to new environmental products. The question is the relative prices of goods like chemicals, food, or electricity and goods like views and health. People may not be able to get the price set and product mix that they want. They may want to act as a group for public goods and change property rights even for individualistic goods, e.g., initial distribution of ownership of water, as various aspects of it become valuable and new uses are envisaged. When tastes change, both demand and supply curves may be shifted.

There are frequently problems in communicating a demand for a nonmarginal change in relative prices and introduction of new goods. In a sense, price sets and the range of alternative goods have some of the characteristics of a public good. They are common to all if they exist at all. Economic analysis needs to present data to help people choose the game they want to play and the broad range of opportunities available as well as how to play a given game to their advantage.

URBAN WATER SUPPLY: A SECOND LOOK

By Jack Hirshleifer, *University of California, Los Angeles*
and J. W. Milliman, *Indiana University*

The two present authors, with James C. DeHaven, of the RAND Cor-
poration, completed in 1960 a study entitled, *Water Supply:Econom-
ics, Technology, and Policy* [5]. Our conclusions had a degree of
novelty and, in some circles, more than a degree of unpopularity. Gov-
ernmental authorities responsible for water-supply decisions, in this
sphere where the private sector is as but a shadow, are tempted to re-
gard themselves not as mere purveyors of a commodity but instead as
crusaders for the cause of ample and pure water. The combination of
crisis and crusade has generally been found effective in enlisting the
support of public-spirited elements of the press, of high-minded citizen
groups, and where necessary of the electorate, for vast new construc-
tions. (It may not be totally amiss to interject that, in the background,
prospects for agency aggrandizement and for beneficiary and vendor
profits also play a role in fueling up crusades.) The economist is likely
to find himself in an awkward position if at a time of crisis he con-
cludes that a particular project, generating an increment of ample and
pure water, is not worth purchasing. Unfortunately, in the absence of
avenues to make criticism effective, water projects have attained of re-
cent years an impressive gigantism that promises to inflict serious eco-
nomic loss upon the communities and regions involved.

I

The *Water Supply* volume discerned a consistent pattern of under-
pricing and overbuilding. Rational reallocation of existing supplies,
such as would occur if prices were raised in response to scarcity, is al-
most never even considered as an alternative to new construction.
There are several reasons, aside from the electoral attractiveness of
water projects: (1) Politically determined prices are inflexible. New
York, for example, has not modified its price schedule since 1933. (Of
course, had the city attempted to do something about prices in the
1965 shortage, they would run into conflict with the CEA "guide-
lines"!) (2) A higher-order political rigidity makes it commonly im-
possible for one water jurisdiction to sell title or rights to another—
hence an incentive to rush into construction to nail down the supplies
involved. (3) Certain errors in economic reasoning have played a role:
ignorance of the marginal principle, double counting of benefits, the

use of inappropriately low discount rates, etc. But these analytical errors have had much less practical significance than what might be called the nonanalytical error. This is the belief, usually quite unconscious, that there are "needs" or "requirements" for water rather than economic demands. No matter how conclusively refuted by observed sensitivity to price (e.g., the all too common experience of inability to sell high-priced water), the belief that demands are absolutely inelastic continues to dominate all planning in this field.

However, it must be conceded that the economist is not yet the possessor of absolute truth. Even where his received theories are adequate, he may still make bad use of them in the world of affairs. We will therefore use this occasion to review the record on two important practical situations—New York and southern California—for which we dared in 1960 to make predictions and recommendations.

II

Our analysis of New York centered about the decision in 1955, taken in consequence of a water crisis five years earlier, to build a large dam (large, that is, by the less gigantic standard of ten years ago) at Cannonsville on the Delaware. In making this decision the New York authorities ignored proposals that would have limited use, or the growth in use, through extension of metering, raising prices, or reducing system leakage. Also, going to the Delaware at Cannonsville was chosen over a nearer connection to the Hudson. Our analysis, as of 1960, supported the contentions that metering and leakage control were highly economical substitutes for new supplies. In addition, we confirmed that as between the alternative new supplies, the Hudson would have been cheaper—though a question about the relative qualities of the two sources remained. As another point of interest, we questioned the solidity of New York's claim to Delaware water—the operative Supreme Court decree having granted New York not a legal right but only a permission subject to reconsideration at any time. Finally, we believed that use projections had been exaggerated even in the absence of the pricing and rationing reforms we thought desirable. We were pleased, therefore, to be able to report in 1960 that New York had decided, after all, to delay completion of Cannonsville to the period 1966-67. Thereby hangs a tale.

Table 1 demonstrates the seemingly overwhelming case for such a "stretch-out" of Cannonsville completion. By 1960 actual use, lagging well behind the forecasts on which the Cannonsville construction had been promised, had stabilized below 1,300 million gallons per day (mgd). Even without Cannonsville, the aggregate "safe yield" of 1,550 mgd provided a very healthy margin over actual and prospective use

TABLE 1

Year	Use* (mgd)	Safe Yield† (mgd)	End-of-year Storage‡ (BG)	Actual Yield§ (mgd)
1955	1,155			
1960	1,258	1,550		
1961	1,285	1,550	250.4	
1962	1,276	1,550	195.6	1,126
1963	1,295	1,550	162.2	1,204
1964	1,262	1,550	124.0	1,157
1965	950–1,000	1,550	181.7	1,133

* SOURCE: Board of Water Supply, *59th Annual Report* (1965), p. 100. (1965 figure estimated.)
† As officially computed. See Northwest Desalting Team [7], pp. 3–7.
‡ *New York Times*, various dates.
§ This is use plus end-of-year increment (or minus decrement) in storage.

reinforced further by a storage capacity of 500 BG—over a year's supply—available for trapping excess runoff of wetter years. In the circumstances, the addition of a supposed 310 mgd of "safe yield" from Cannonsville seemed highly postponable.

What went wrong is evident in Table 2. The year 1960 was the last of "normal" rainfall in the New York area. In the phenomenal drought that followed, actual yields fell not only below safe yields but below actual use—requiring draw-down of accumulated storage. Emergency restrictions on water use were imposed in April, 1965, and succeeded in reducing the rate of utilization some 25 percent to around 950 mgd. This reduction also reflected campaigns for voluntary conservation and efforts to stem leaks throughout the system. Nevertheless, by mid-1965 crisis was at hand; the near prospect was for empty reservoirs and dry faucets. In June, New York unilaterally stopped releasing water downstream as required under the Supreme Court decree entitling it to use of Delaware water. This action in turn threatened saline contamina-

TABLE 2

RAINFALL STATISTICS NEW YORK 1959–65*

Year	Rainfall in Inches	Percent Above or Below Average (42″)	Percent Accumulated Deficiency
1959	38.77	− 7.6	
1960	46.39	+10.4	
1961	39.32	− 6.4	6.4
1962	37.15	−11.6	18.0
1963	34.28	−18.4	36.4
1964	32.99	−21.5	57.9
1965	26.09	−37.9	95.8

* SOURCE: Raymond J. Faust. "Northeast Water Crisis and its Solution," *J. American Water Works Association*, Jan., 1966. (1965 figures from *New York Times*.)

tion of the supplies of Philadelphia and other cities tapping the Delaware downstream. A series of compromises was worked out by the Delaware River Basin Commission, in effect requiring just enough releases from New York to keep the salt-water front safely below the Philadelphia intake. Private power companies contributed substantially by releasing storages on tributaries of the Delaware. Nevertheless, the situation was rescued only by good rains in the last quarter of 1965. (See [3] [4] [9] [10].)

What would have been the effect in the 1965 crisis of earlier construction of Cannonsville, alternatively of a Hudson supply? It has been maintained that Cannonsville would have saved the day. Cannonsville could at most have stored around 50 BG more than it already held in 1965—about fifty days' supply at the reduced rates. But, since much higher releases downstream to the Delaware would have been legally required of New York upon completion of Cannonsville, it seems highly doubtful that the additional accumulation could or would have occurred. In fact, the required releases are so great that Cannonsville can hardly be said to provide any increment of yield to the city in time of drought. On the other hand, water in the Hudson in very adequate volume was flowing uncontrolled past New York during the entire period. (New York's actual use of the Hudson has an unfortunate history. An emergency connection to yield 100 mgd was rushed to completion in 1950, but arrived too late to assist in the crisis of that year. Subsequently, it was completely dismantled, only to be reconstructed as an emergency project again for the 1965 crisis. Again too late, it was shut down after pumping only seven days.)

What can we learn from the experience briefly reviewed here? First of all, perhaps, would be a healthy skepticism toward meteorology and hydrology. Here we have a case where predictions as to the social phenomena of use, while off the mark, still proved far more reliable than predictions as to the natural phenomena of supply. Analytically, we must take better account of uncertainty and variability of supply. In particular, safe yield must not be treated as a certainty-equivalent. In inferring from the evidence of the low rates of return on water investment that systems were typically overbuilt, we failed to allow fully for the insurance aspect of extra supplies. On the other hand, variability makes all the stronger our argument as to the necessity for making better use of existing supplies as an alternative to new construction yielding uncertain water increments. A related consideration is that the variability of alternative new sources should be considered in weighing investment options. In the case of New York, the Hudson would have been a more reliable source than the Delaware.

The legal history of New York's use of Delaware water is also significant. New York, in return for temporary permission to withdraw

some 490 mgd from the Delaware, committed itself to downstream re-leases adequate to maintain a flow of 1,525 cfs at Montague, N. J. In the dire emergency New York was able to pressure the other states, despite the danger to such cities as Trenton and Philadelphia, to con-sent to a relaxation. Evidently, in time of crisis the agency in actual possession or control of the upstream source is in a strong position re-gardless of legal technicalities.

We may terminate this section with some comments on the present prospects for New York.[1] The immediate situation looks somewhat better because of more normal rainfall. The safe yields have been writ-ten down some 20 percent, but still inspire little confidence. Rationali-zation of existing uses has made a little progress. Efforts during the crisis to detect and repair leaks seem to have tightened the system con-siderably, though controversy persists on this point. Recirculation de-vices and conservation equipment have been required on industrial cooling and on air-conditioning units. Reluctance to use the price mechanism is as strong as ever, however; an increase in the water price seems out of question. And while the principle of extension of metering is widely accepted, actual progress in that direction seems stalled. Cannonsville will be in full operation shortly—and the Hudson emergency tap provides a small reserve. But overall, New York re-mains very vulnerable in the absence of a return to long-term "nor-mal" rainfall patterns.

The 1965 crisis also brought forward the usual desalinization en-thusiasts. In this case the Great Society rushed into the act, on direct instructions from President Johnson, and an interagency Northeast Desalting Team was appointed. Reputedly, economists on the Team played a creditable role in insisting upon comparison with alternatives to desalinization—a process which showed that a Hudson source would cost only half as much [7]. The inconvenience of having economists around was thus once more demonstrated to the grand projectors dom-inant in the water supply field.

III

The water experience of southern California[2] since 1960 provides an interesting contrast with New York, accentuated by the similarity of the continuing drought in both areas.[3] In New York a seeming surplus in terms of safe yields was revealed by hard experience to be illusory, and the supply remains tight. While new construction has taken place,

---

[1] For an excellent review, see the Citizens Budget Commission report [2].

[2] More technically, the South Coastal Area. This includes the coastal plains of the popu-lous southern counties, with some 10,600,000 inhabitants (1965)—over half the population of California.

[3] In only four of the past twenty years has rainfall in southern California been "average" or better.

for hydrological and legal reasons the city has reaped little or no benefit. In California, on the other hand, a political "water coalition" has created what appears to be an already excessive and costly supply; the coalition continues to seek out vast new increments, although supplies now arranged for cover growth needs for thirty to fifty years. In neither area has the responsible authorities shown much interest in rationalizing existing uses by appropriate pricing and reallocation.

In 1960 the voters of California, by a narrow margin, endorsed an enormous bond issue ($1,750,000,000) covering a first installment on the Feather River Project—the central component of the state's "Water Plan." We argued that the project was premature, that an increase in water rates should have come first, and that cheaper alternatives were available from reallocation of agricultural supplies and from reclamation of waste waters. Throughout southern California water has been priced below the marginal cost of existing supply, not to mention the higher-cost Feather River supply. Discriminatory pricing in favor of agriculture and failure to reallocate water from agriculture to the growing urban sector have led to wide discrepancies in marginal values in use—within the area and also between the South Coastal Area and nearby agricultural regions (especially Imperial Valley).

Developments since 1960 are summarized in the source-use balance of Table 3. Lack of space prevents detailed review, but in the half decade water use rose 13 percent, somewhat less than the remarkable population increase of 22 percent. (Adjusting for spreading to replace underground withdrawals, water consumption rose just 9 percent.) The divergence is explained by the decline in agricultural use, which nevertheless remains some 25 percent of the total. Discriminatory pricing has encouraged bringing outlying acreage under irrigation as subdivisions have displaced existing agriculture. Even at current relatively lavish consumption rates (160 gallons per capita per day), this agricultural water would support a further urban development of some 3,500,000 people—probably a decade of growth. And more rational pricing to residential and industrial users would easily extend the urban supplies another half decade.[4]

Setting aside the rationalization of existing uses, Table 4 summarizes the existing and prospective sources. The margin of present development over the current rate of use is relatively small, but the huge Feather River supply coming in 1972 will assure that water will always be physically available, into the foreseeable future. But economic availability is another question.

The Metropolitan Water District—water wholesaler to the area—

_____
[4] Urban demand elasticities are around .4; a 50 percent price increase, certainly called for, would cut use around 20 percent.

TABLE 3*

Source-Use Balance for South Coastal Area

	Withdrawals for Use (Acre-Feet)		
	1950	1960	1965
Water Source:			
Local surface and ground water			
1. Developed safe yield	1,066,000	1,098,000	1,098,000
2. Gross overdraft†	(300,000)	(300,000)	(300,000)
3. Net overdraft†	275,000	88,000	(20,000)
Los Angeles Aqueduct	320,000	320,000	320,000
Colorado River Aqueduct			
(capacity 1,200,000 acre-feet)			
1. Agricultural and urban use	142,000	523,000	796,000
2. Water spreading	25,000	212,000	320,000
Total	1,828,000	2,241,000	2,534,000
Water Use:			
Irrigation	954,000	700,000	630,000
Urban	874,000	1,541,000	1,904,000
Total	1,828,000	2,241,000	2,534,000
Water-Use Assumptions:			
1. Agricultural use			
Irrigated acres	650,000	500,000	450,000
Water applied in feet	1.5	1.4	1.4
2. Urban use			
Population	5,480,000	8,656,000	10,600,000
Use rate in gallons per capita per day .	140	160	160

* This table is an extension of Table 34 in *Water Supply*, p. 295.

† Gross overdraft refers to total pumping in excess of long-run recharge. Net overdraft is gross overdraft minus the spreading of imported waters. In 1965 the spreading exceeded gross overdraft, thereby overstating actual water consumption by 20,000 acre-feet.

was pressured into a commitment to absorb the enormous quantities to be delivered by the state; the price is likely to be very steep, around $65 per acre-foot of raw water. MWD must also cover the large expense of treatment and distribution. To meet these heavy commitments, the MWD established a policy in 1964 of setting rates high enough to bring in revenues sufficient to cover all operations costs plus one-half of all capital costs. From 1960 through 1965 wholesale prices for municipal water were doubled while rates for agricultural and replenishment use were increased about 30 percent. Planned increases through 1975 would raise prices for municipal use to $58 an acre-foot, up from $15 in 1960, and for agricultural use to $26 an acre-foot, up from $12 in 1960.[5] (See [11] [12].) However, we believe that there is little hope of selling much water at these prices. The MWD, thus, may have to load much of the financial burden on the tax base, just as it has done in the first thirty years of its existence. Paradoxically, the planned water rates for the 1970's will very likely be too high; once

[5] A surcharge of $9.00 per acre-foot should be added for filtered and softened water.

the facilities for the Feather River supply are in place the marginal costs of supply will be relatively low. If we are correct that the Feather River supply will remain largely unsalable for the next several decades (if planned pricing policies are followed), this will repeat the deplorable history of the long-unused Colorado River aqueduct—another monument to the view that there are absolute "requirements" for water rather than demands at a price. (See [6].)

We may comment briefly on the remaining items of Table 4. The second barrel for the Los Angeles Aqueduct is an economical project, in the range of $24-$30 per acre-foot at 1965 costs [8]. This is a city of Los Angeles undertaking; being cheaper than even the out-of-pocket costs of buying MWD water, its construction was rational for the city despite the incoming Feather supply. What is scarcely comprehensible, however, is why Los Angeles consistently pushes the MWD into commitments for tremendous costly supplies, after which the city

TABLE 4

SOURCES OF WATER SUPPLY FOR THE SOUTH COASTAL AREA

	Acre-Feet
A. *Present Development:*	
1. Local surface and ground waters	1,098,000
2. Los Angeles Aqueduct	320,000
3. MWD's Colorado River Aqueduct	1,200,000
	2,618,000
B. *Scheduled Development:*	
1. Second Barrel to Los Angeles Aqueduct (1968)	150,000
2. Feather River Water (1972)	2,000,000
3. Nuclear Desalting Plant (1972–80)	150,000
	2,300,000
C. *Scheduled Loss:*	
MWD rights to Colorado River water expected to be reduced by 650,000 acre-feet per year on a gradual basis from 1975–90. Rights in 1990 to total 550,000	−650,000
Subtotal	4,268,000
(Developed or scheduled for 1990)	
D. *Potential Sources of Supply at Low Cost:*	
1. Reclamation of sewage and waste waters	450,000
2. Purchase of Colorado River rights from Imperial Irrigation District to fill unused capacity in Colorado River Aqueduct	650,000
	1,100,000
Grand Total	5,368,000
(Developed, scheduled, and low-cost supplies)	

E. *Higher Cost Water Sources:*
1. Purchase of additional Colorado River water for transfer to the area.
2. Additional reclamation of sewage and waste waters.
3. Purchase and transfer of water from the San Joaquin Valley to the area.
4. Additional transfer of water from northern California.
5. Additional desalting plants.
6. Conveyance from Columbia River, Canada, Alaska, and the Moon.

builds and uses its own cheaper sources instead. Since one-third (formerly one-half) the MWD tax burden falls on Los Angeles, the cost to its taxpayers has been enormous.[6]

The prospective loss of MWD entitlement to 650,000 acre-feet of Colorado water, as a result of the Supreme Court decision favoring Arizona's claims, constituted one of the stronger arguments for the Feather River Project. Even before this decision, however, we pointed out that purchase of a portion of the 3,850,000 acre-feet of prior agricultural entitlements within California to Colorado water (from Imperial Valley and certain other districts), for transport to the South Coastal Area, would have been cheaper than the Feather supply. The Supreme Court decision will have the effect of leaving MWD's Colorado River aqueduct half-unused again (when Arizona fully exploits her established rights); hence, 650,000 acre-feet of purchased agricultural entitlements could be transported via existing, otherwise idle capacity. Finally, reclamation of waste waters would have been another cheap source, far cheaper than the Feather supply. Of course, in the absence of escape from the Feather River commitment, MWD's problem at this point is disposing of water rather than seeking out new supplies, however economical.

Astonishingly, however, the MWD has gone ahead and committed itself further to a new source: nuclear desalinization. This was justified in a Bechtel Corporation study [1] by the consideration that the water (at about $88 per acre-foot) was cheaper than conventional delivery from the Columbia. (Next, presumably, a Columbia project will be justified as cheaper than delivery from sources in Canada and Alaska; not entirely a joke, as such a project is indeed being promoted!) Since this desalinization is a tie-in deal with electric power (the water production, at about 150,000 acre-feet, being relatively modest), it may be hoped that the power economics are more attractive. The press-agentry values of the project, in terms of enthusiastic newspaper reports, were considerable—thus possibly reducing resistance to the tax increases MWD will have to impose for paying off its Feather commitment.

IV

It is difficult to avoid a critical tone in reviewing what appear to be instances of inefficient stewardship of water resources by political or bureaucratic authorities. Certainly they ought not go on committing

[6] The city has paid some $250 in MWD taxes for every acre-foot used (in comparison, the actual water price charged by MWD has been negligible). San Diego has, on the other hand, paid only 8 percent of the taxes while taking 25 percent of the water; its tax burden has only been $20 per acre-foot.

the same errors, having been deluged by a river of advice from econo-mists, not only ourselves but also from Eckstein, Krutilla, McKean, Maass *et al.*, Renshaw, Kneese, Haveman—most recently, we can wel-come to the club the new work of Bain, Caves, and Margolis. But is our critical tone justified? Decision-makers are just people, and people will make mistakes. In the private sector, those who regularly commit mistakes lose control, in the long run, over the disposition of resources. In the government sector, this process operates weakly, if at all. It ap-pears that the agenda for economists, at this point, should place lower priority upon the further refinement of advice for those efficient and selfless administrators who may exist in never-never land. Rather, it should center upon the devising of institutions whereby fallible and imperfect administrators may be forced to learn from error. In short, can we construct a "hidden hand" for the government sector?

REFERENCES

1. *Engineering and Economic Feasibility Study for a Combination Nuclear Power and Desalting Plant* (Bechtel Corp., Dec., 1965).
2. Citizens Budget Commissions, Inc., "Action on Water" (Nov., 1965).
3. Delaware River Basin Commission, *Annual Report*, 1965.
4. "How to Run a River," *Federal Reserve Bank of Philadelphia Business Review*, Sept., 1965.
5. J. Hirshleifer, James C. DeHaven, and Jerome W. Milliman, *Water Supply: Eco-nomics, Technology, and Policy* (Univ. of Chicago Press, 1960).
6. Jerome W. Milliman, "The History, Organization, and Economic Problems of the Metropolitan Water District of Southern California" (U.C.L.A. Ph.D. thesis, 1956).
7. Northeast Desalting Team, "Potentialities and Possibilities of Desalting for Northern New Jersey and New York City," June, 1966.
8. Max K. Socha, "Construction of the Second Los Angeles Aqueduct," *J. of American Water Works Asso.*, June, 1965.
9. Water Resources Association of the Delaware River Basin, *Delaware Basin Bul.*, Vol. 6 (Sept., 1965).
10. James F. Wright, "Water Resources of the Delaware River Estuary," *J. American Water Works Asso.*, July, 1966.
11. Metropolitan Water District of Southern California, *27th Annual Report* (1965).
12. Metropolitan Water District of Southern California, *Report No. 821: 1965 Water Pricing Investigation* (Apr., 1965).

WATER POLICY AND ECONOMIC OPTIMIZING: SOME CONCEPTUAL PROBLEMS IN WATER RESEARCH*

By S. V. CIRIACY-WANTRUP
University of California, Berkeley

I. *Water Policy and Water-Resources Systems*

For the purpose of this discussion, I should like to view water policy from the standpoint of decision theory: water policy may be regarded as a set of decision rules in a multistage decision process. In this process, a sequence of decisions extends over time and space in an "open" system.

The first step in such a study of water policy[1] is to identify the system or systems in the control of which decision rules are sought. In other words, what are the characteristics of the water-resources system with which water policy is concerned? This seemingly simple issue of system identification is, I submit, at the root of some of the most serious conceptual problems in contemporary water research.

The large and still expanding literature on water economics during the last decade is mainly concerned with one particular class of a water-resources system; namely, the public multipurpose development of surface water through storage dams, canals, and other large engineering structures. The term "public" usually refers to the federal government, and economic analysis focuses on the efficiency of federal investment in water resources. This is the same class with which the earlier discussion of benefit-cost analysis was concerned—exemplified by the Green Book, the controversy about the TVA, and the various critiques of the Army Engineers and the Bureau of Reclamation. The Green Book used the term river-basin project for this class, but later the terms water-resources system and water development became synonymous with it.

Such terminology is not helpful for several reasons. First, this particular class is less significant in inputs and outputs than other classes. Second, this class is always closely related to others as a part of an integrated system; it is merely a subsystem. Third, the decision rules suggested in the literature for this subsystem under constraining assumptions are neither valid nor relevant for the integrated system as a

---

* This will be published as Giannini Foundation Paper No. 272.

[1] In the terminology of decision theory, policy refers to the decision rules; that is, the criteria of decision making. But the study of policy includes the whole decision-making process and the implementation and the effects of decisions.

whole. The last point will be elaborated presently (Section II), but first let us ask what are the characteristics of an integrated water-resources system as understood here?

Such a system is a mixed groundwater-surface water system. Groundwater use is quantitatively at least as significant as surface-water use, and integration of the two uses raises some of the most important issues for water policy. The design of such a system does not necessarily involve large engineering structures. Appropriate institutional structures, on the other hand, are a necessary and frequently sufficient condition for its functioning. Such institutions relate to water law influencing water development, water allocation, and water quality; to water district law controlling the establishment, organization, and operation of public water districts; and to state and federal administrative agencies affecting water development, allocation, and quality. For short, these structures will be called here "water institutions."

Groundwater is developed largely by private rather than by public investment. The public investment that is involved is by water districts rather than by federal and state governments. Private firms and water districts are also active in multipurpose, multiunit surface-water development. Some of the federal developments and the California State Water Plan outrank private and district developments in size of individual projects; but in the aggregate, water development by private firms and water districts exceeds that by federal and state governments.

We may say, then, that the water-resources system with which we are concerned consists of operating sectors which are private firms, public districts, and projects of federal and state governments.[2] In aggregate quantity of water developed, private firms—households, farms, industrial corporations, and public utilities—are the most significant group of operating sectors. In second place are public water districts such as irrigation districts, municipal water districts (and departments), and conservancy districts. Federal projects are in third place and state projects are last.[3]

[2] It should be noted that a differentiation is made here between projects of federal and state agencies and those agencies themselves. This is in accordance with procedures of economic optimizing. Thus, optimizing procedures are applied to projects of the Bureau of Reclamation and not to the Bureau itself or the laws regulating its establishment, organization, and operation. A parallel differentiation is made here between public water districts and the state water district laws controlling the establishment, organization, and operation of such districts.

[3] In recent decades, water districts have changed in part from water developers in their own right to retailers of water developed by federal and state projects. In some states—for example, California—federal and state projects are intertwined. Statistical separation of quantities of water developed by districts, federal projects, and state projects has become increasingly difficult. For a more detailed discussion of the relative significance of the vari-

II. *Water Policy and the Hierarchy of Decision (Optimizing) Levels*

In order to find decision rules for such an integrated water-resources system I should like to differentiate three levels of decision making. These levels constitute a logical pyramid similar to the levels of conceptualization differentiated in formal logic.[4] For the present purpose, this pyramid will be called "the hierarchy of decision levels."

On the first level, the lowest, the decision-making process relates directly to the control of inputs, outputs, and other quantitative characeristics of the water-resources system. These characteristics may be deterministic or stochastic; they may be in physical or in value terms. Decision making on the next higher level, the second, controls the institutional framework of the decision-making process on the first level. On the third level, the framework of the decision-making process on the second level is the subject of decisions.

Decision levels may also be conceived as optimizing levels. This term will not be used here because it requires interpretation of optimizing as a fictional construct or scientific fiction.[5] On each level, decision rules are sought for making the best decision on that level. Although decision-making processes differ from level to level, they are interrelated because the effects of each decision can be traced through all lower levels. From the observation of these effects on the water-resources system, decision-makers can learn how to make improvements in decision-making processes on all levels. Such learning—or in computer jargon "feedback"—is the essence of a multistage decision process. The decision-making process on each level can be studied in its structure, its functioning, and its performance.

We can now proceed to the application of these general concepts. As we know, an integrated water-resources system is composed of private and public operating sectors (Section I). The decision making on the level of these operating sectors constitutes the first level of our hierarchy. Decision rules for this level are familiar to economists because they are identical with "the" rule generally suggested for decision making in economic literature. This rule was originally postulated for a calculative economic man and for the private profit-seeking firm. It

ous groups of operating sectors in the above sense, see S. V. Ciriacy-Wantrup, "Water Policy," *Handbook of Applied Hydrology: A Compendium of Water-resources Technology,* Editor-in-Chief, Ven Te Chow (McGraw-Hill Book Co., 1964), Sec. 28, pp. 28-1 to 28-25.

[4] For an application of levels of conceptualization to the concept: "resources," see Ciriacy-Wantrup, *Resource Conservation: Economics and Policies* (2d ed. rev., Div. of Agric. Sci., Univ. of California, 1963), Chap. 3. It scarcely needs to be mentioned that the logical pyramid referred to here carries no connotation of social or ethical ranking.

[5] For a discussion of this interpretation, see Ciriacy-Wantrup, "Policy Consideration in Farm Management Research in the Decade Ahead," *J. of Farm Econ.,* Dec., 1956, pp. 1301-11.

was later modified for public decision making in order to take account of problems involved in aggregating individual utilities and in dealing with externalities. The recent revival of the concept of social time preference may be regarded as a special case of these two problems. The public maximizing agent striving to obtain such Pareto optima in water-resources development by internalizing externalities has been dubbed the "river basin firm."[6]

The common decision rule for private and public operating sectors specifies maximization of an objective (profit or welfare) function under constraints regarding institutions, technology, and resource availability. In terms of formal criteria, maximization is accomplished through fulfilling the necessary and sufficient conditions and other qualifications given by the first and second derivatives of the objective function and, for maximization over time, by the calculus of variations.

Little needs to be said here about this decision rule itself. The conceptual and operational limitations of quantitative optimizing in private and in public decision making have been discussed elsewhere.[7] There is, however, one aspect of these limitations that is of interest here because it points to the next higher decision level. This aspect is the treatment of institutions as constraints in quantitative optimizing. Such treatment does not create conceptual problems in private decision making. In social decision making, however, institutions correspond conceptually more to independent or dependent variables than to constraints.

The latter treatment of institutions has received critical review in welfare economics through "the theory of second best."[8] This critique has cast serious doubts on the validity of Pareto optimizing under institutional constraints. Whether or not one agrees with the entire reasoning of this critique, the main result can well be accepted: institutional conditions are conceptually different from technological conditions and resource availability. To look at the former as constraints views their appearance only on the first decision level. On the second level, they appear as the operational parts in the decision-making process.

[6] Allen V. Kneese, *The Economics of Regional Water Quality Management* (Johns Hopkins Press, 1964).

[7] Ciriacy-Wantrup, "Philosophy and Objectives of Watershed Policy," *Economics of Watershed Planning,* ed. G. S. Tolley and F. E. Riggs (Iowa State Univ. Press, 1961), pp. 1-12; and "Conservation and Resource Programming," *Land Econ.,* May, 1961, pp. 105-11.

[8] R. G. Lipsey and R. K. Lancaster, "The General Theory of Second Best," *Rev. of Econ. Studies,* Vol. XXIV(1), No. 63 (1956-57), pp. 11-32; John V. Krutilla, "Welfare Aspects of Benefit-Cost Analysis," *Economics and Public Policy in Water Resource Development,* ed. Stephen C. Smith and Emery N. Castle (Iowa State Univ. Press, 1964), pp. 22-33; and O. A. Davis and A. B. Whinston, "Welfare Economics and the Theory of Second Best," *Rev. of Econ. Studies,* Jan., 1965, pp. 1-14.

The purpose of decision making on the second level is not to control directly inputs, outputs, and other quantitative characteristics of the water-resources system nor to obtain a path of quantitative welfare optima at various points in time under projected conditions for these points. Rather, the purpose is to maintain and to increase welfare by continuously influencing decision making on the lower level under constantly changing conditions that for any point in time cannot be projected—or only vaguely—and that are always uncertain with respect to actual occurrence.

In order to fulfill such a purpose, each water institution may be regarded as a decision-making system that functions as a whole with a particular pattern of change. Under the constitutional organization of the United States and other Western democracies, the system is modified through actions by the three branches of government—the legislative, judicial, and executive—each with a different range over which modification can be accomplished. Modification takes place on the federal, the state, and the local levels; generally the state level is the most important for water institutions.

Performance of such a decision-making system can be appraised only by viewing the system as it functions over time under various economic conditions. It is conceptually inadequate to appraise performance by studying temporal cross-sections of the system for particular conditions and points in time. Criteria for performance must be sought on the system's own level. They need not be the same as those on the lower level discussed previously. Neither need the criteria on the lower level be those that economists suggest they should be. What we seek for the second level are criteria that could serve as conceptually and operationally meaningful proxies for the fictional construct of optimizing welfare. They may be called intervening criteria in analogy with intervening variables. The search for such criteria will be undertaken in the following section (Section III).

While the second decision level is the most significant one for the study of water policy, it does not complete our hierarchy. A third decision level was implied by the reference to the constitutional organization of the United States, which sets the basic framework for water policy as for all other policies. Water institutions, however, are not confined to modern Western democracies. Highly developed water institutions existed in ancient feudal societies and city states in the Old and in the New World. They exist in primitive tribal communities in Africa no less than in modern authoritarian states. Time does not permit sketching the differences between water institutions in different societies. Although the causal connection is controversial, the basic organization of a society is always closely interrelated with water institutions. Accordingly, this organization may be regarded conceptually as

a social decision-making process on the next higher, the third, level of our hierarchy. Discussion of this level would lead away from water policy and need not be undertaken here.

III. *Criteria for the Performance of Water Institutions*

Decision-making processes on the second and third levels are a part of the political process. When economics was understood as political economy, it encompassed the political process in the study of decision making. Later, emphasis shifted toward the quantitative optimizing model almost to the exclusion of other decision-making processes. However, there are some nonconformists. For example, Lindblom and his coauthors in several books, especially in the last one, argue that the political process is the only valid and relevant general model of decision making.[9] He applies his central theme of "mutual accommodation of partisans" to private as well as to public decision making. Others have suggested that the political process should merely supplement quantitative optimizing in the solution of specific technical difficulties; for example, to determine trade-off values between different objectives of public investment.[10] Multidimensionality of the objective function has long been a major conceptual and operational difficulty in quantitative optimizing.

These and similar suggestions are a refreshing change from reiterating over and over the goal of quantitative optimizing—often with little regard for the operational possibility of fulfilling the necessary and sufficient conditions and other qualifications. There are, however, two provisos which I should like to make: First, one must recognize the differences between decision-making processes on different levels: the political process must be brought in at the second level; it cannot be relied upon to solve *ad hoc* the many technical difficulties of quantitative optimizing. Second, conceptually satisfactory criteria must be provided to differentiate a "good" from a "bad" decision.

By the authors mentioned and by others, agreement based on mutual accommodation of partisans is specified as the main criterion. Such a criterion has strong appeal in a Western democracy. But its application involves several conceptual difficulties. The criterion requires a careful definition of the meaning of political agreement and a

---

[9] Robert A. Dahl and Charles E. Lindblom, *Politics, Economics, and Welfare: Planning and Politico-Economic Systems Resolved into Basic Social Processes* (Harper, 1953). David Braybrooke and Charles E. Lindblom, *A Strategy of Decision: Policy Evaluation as a Social Process* (London: The Free Press of Glencoe, Collier-Macmillan Limited, 1963). Charles E. Lindblom, *The Intelligence of Democracy: Decision Making Through Mutual Adjustment* (The Free Press, 1965).

[10] Arthur Maass, "Benefit-Cost Analysis: Its Relevance to Public Investment Decisions," *Q.J.E.*, 1966, pp. 208-26.

specification of the means of bringing it about. While the criterion is operational in the sense that "feasible" decisions are always selected, it does not enable one to make a selection between alternative decisions for all of which agreement is attainable. Agreement may fluctuate over relatively short periods of time for substantially identical decisions; this would not make the criterion inoperative at any point in time, but differentiation between good and bad decisions would have little meaning under these conditions. Finally, and most importantly, the agreement criterion is not useful in appraising performance for the purpose of scientific analysis.

To find a more satisfactory criterion is not easy. But since the social sciences may be regarded as an extension of human ecology, performance criteria used in our sister sciences may offer a suggestion.

Geneticists, studying genotypes, populations, and species, differentiate between favorable and unfavorable gene variants, mutations, traits, and other characteristics.[11] Students of animal behavior differentiate between favorable and unfavorable instincts and other habit patterns of behavior with various degrees of openness to learning.[12] These and similar appraisals of characteristics—from molecular to social—are based on a common criterion; namely, survival under the pressure of selection. One might explore the possibility of applying a similar criterion to characteristics of economic behavior, in our case to the performance of water institutions. Here, survival must be interpreted in economic terms; that is, not in physical growth and numbers but in economic growth and welfare.

The first step in such an exploration is to note some relevant conceptual implications of survival value as a criterion of performance. Survival value indicates only direction, even though highly quantitative methods may be used in determining it; it cannot be employed, therefore, for obtaining optima. It has no normative connotation; it is useful for scientific analysis; but for political decision making, its usefulness is indirect and uncertain. Knowledge of structure and functioning of a system is required before performance of its individual characteristics can be appraised. Appraisal is valid only for a specified environment or sequence of environments.

None of these implications render the criterion unsuited for our purpose. It will be recalled that quantitative precision and optimizing are not involved on the second level of decision making. Further, a criterion is desired that is useful for scientific analysis. For political deci-

[11] Curt Stern, "The Genetic Resources of Man," *Natural Resources: Quality and Quantity*, ed. S. V. Ciriacy-Wantrup and James J. Parsons (Berkeley: Univ. of California Press, 1967) (in press). Ernest Mayr, *Animal Species and Evolution* (Harvard Univ. Press, 1963).
[12] Konrad Lorenz, *On Aggression* (Harcourt, Brace and World, 1963). N. Tinbergen, *The Study of Instinct* (Oxford: Clarendon Press, 1951).

sion making, such a criterion can only be a supplementary one, its significance depending on the influence of scientific understanding on political agreement.

Survival value can be applied in various ways. Sometimes it is the performance of water institutions as it affects the viability of operating sectors (in the sense of Section I) that is of interest; for example, how water-district law affects the viability of water districts during economic depressions or their growth in underdeveloped regions. Sometimes it is the water-resources system itself that is of interest; for example, to what extent are water institutions responsible for the growth, stagnation, and decay of irrigation systems in many parts of the world? Usually, however, scientific interest is focused on how water institutions—especially water law—are related to the welfare of a whole region in various periods of its development.

IV. *Illustrations and Conclusions*

In order to bring the conceptual analysis (Sections I-III) closer to applied water research, some of these examples may be spelled out further.

Among the most difficult problems of water policy is the allocation of costs of water development among beneficiaries for the purpose of repayment. For water districts, solution of this problem is crucial for viability because they cannot rely on the resources of the federal and state treasuries. Water research can use two approaches: First, through various computational techniques of benefit-cost analysis one can project benefits accruing to different groups of beneficiaries and allocate costs and repayment more or less in proportion. The second approach focuses on alternative characteristics of water-district law. The objective of this approach is to appraise the allocative performance of these characteristics and to select those with survival value in terms of viability of districts.

Since the Wright Act of 1887, California has had a great deal of experience in employing district taxation and water prices in various combinations. With some modification in 1907, the original taxation provisions of the Wright Act have stood up well under the test of cyclical and structural economic change. Conversely, the difficulties of water districts in other states, in Canada, and elsewhere are largely due to the absence of similar provisions in water-district law. The details are of great interest for institutional analysis, but time does not permit discussing them here.[13] In summary, one may suggest that as a

---

[13] For details, see Michael F. Brewer, *Water Pricing and Allocation With Particular Reference to California Irrigation Districts* (Univ. of California, Giannini Found. Mimeo. Report No. 235, 1960).

basis for repayment to water districts, computation of indirect benefits and of optimum cost allocation has yet to demonstrate their superiority over the results of institutional characteristics tested over some eighty years. In other words, a decision-making process on the second level, by controlling decision making on the first level, has performed in a way that is conceptually and operationally more acceptable than *ad hoc* sophisticated computations. One might call the taxation provisions of the Wright Act an institutional characteristic with survival value.

By far the most significant of water institutions are the systems of water rights that function in great variety in the fifty states and in most organized communities all over the world. Here, also, California offers interesting material for appraising performance under a sequence of different economic environments.

Starting in the 1850's, when industrial use of water—for placer mining—was more important than agricultural use in a pastoral, nonurban region, water rights have controlled water development, allocated water, and influenced water quality up to the present day, when agricultural use of water is dominant in one of the most highly industrialized states of the union.[14] This adaptability of water law in responding to and influencing a rapidly changing economic environment can be explained largely by the functioning side by side of two legal systems, one based on riparian, the other on appropriative rights. Over time, they have come to a workable blend. For groundwater, this blend can be regarded as a specific system, that of correlative rights.[15]

To the superficial observer, the California "mix" of water-right systems appears as utter confusion. It is often criticized by economists who have made it responsible for retardation of water development, misallocation of water, cross-hauling of water, and other inefficiencies. Instances of this kind exist, although other factors are frequently more responsible for them than water law. Even if water law were solely responsible, economic inefficiencies under particular conditions at particular points in time are not sufficient for an indictment of the whole in-

---

[14] For comparing different uses quantitatively, two factors are frequently not sufficiently considered: (1) whether conveyance losses are included or excluded for agricultural use and (2) whether water use by steam-electric plants—the quantitatively most significant one among industrial uses—is included or excluded for industrial use. In California, for example, agricultural use is 87 percent; industrial use, 5 percent; and domestic use, 8 percent of total use, if conveyance losses are included and steam-electric plants are excluded. Agricultural use is 67 percent; industrial use, 25 percent; and domestic, 8 percent, if conveyance losses are excluded and steam-electric plants are included. In terms of water consumption, the former is a more appropriate comparison, provided that double counting is avoided. Some conveyance losses are used via groundwater and counted then. There is considerable interfirm reuse of water both in agricultural and in industrial use; the quantitative extent of such reuse is not known. Steam-electric plants use, largely, cooling water that is not usable for other purposes and, in any event, is not consumed. Potentially, of course, most domestic use can be made nonconsumptive.

[15] Ciriacy-Wantrup, "Some Economic Issues in Water Rights," *J. of Farm Econ.*, Dec., 1955, pp. 875-85.

stitution. As we have seen (Section II), the function of water institutions is not to maximize economic efficiency for particular conditions and points in time but to structure decision making on the lower level under various and constantly changing conditions.

In studying the performance of water-right systems in fulfilling this function, a dichotomy of criteria has been developed.[16] They are security against legal, physical, and tenure uncertainties and flexibility in various legal and economic categories. Applying these criteria to California water law gives a fairly good "fit." California water law has performed relatively well—as compared with other water-law systems —in stimulating water development, allocating water, and protecting water quality over more than a century of profound changes in the economic environment. To be sure, some structural characteristics, such as the system of preferences, have lost survival value and are largely neutralized by others.

The purpose of such a general evaluation is not to defend California water law but to suggest that in the perpetual process of legal adaptation, economists could be more helpful by making careful institutional analyses of the performance of water law than by reiterating the criticism that water law has failed to optimize water development and allocation for particular conditions and points in time. Such criticism is neither valid nor relevant because its criteria are not applicable to a system operating on the second decision level.

If I may draw a conclusion for water research, it is that comparative analysis of water institutions is a promising field for economists. Like all institutional analysis, it is closely related to the political preferences, the emotions, and the social conditioning of the investigator— witness what some investigators have called the "rape" of Owens Valley. This challenge to scientific attitude must be faced. Even though some material is slanted and some merely descriptive, much is available in economic history, political science, law and engineering that is valuable for an analytical treatment of the structure, the functioning, and the performance of water institutions. In this treatment, theoretical constructs and their testing are no less needed than in the analysis of the marketplace. The hierarchy of decision levels is the most important of such constructs. From it follows the basic difference between decision rules (criteria) on the first and second levels.

As to criteria for decisions on the second level, the approach suggested here focuses on what we called intervening criteria as proxies

[16] Ciriacy-Wantrup, "Concepts Used as Economic Criteria for a System of Water Rights," *Economics and Public Policy in Water Resource Development*, ed. Stephen C. Smith and Emery N. Castle (Iowa State Univ. Press, 1964), pp. 251-71; and "Water Economics: Relations to Law and Policy," *Waters and Water Law in the United States*, ed. Robert Emmet Clark (Chicago: Allen Smith and Co., 1967) (in press).

for optimizing welfare; namely, on institutional characteristics that have demonstrated favorable or unfavorable effects on welfare. Examples were given for two of the most important water institutions. Other institutions would lend themselves to the same approach.

This approach is greatly interested in economic history and in relating one time period to another but not necessarily through increasing the number of variables and equations in mathematical models. It relies heavily on theoretical constructs but less on the maximization principle. Emphasis is on determining conditions for economic growth rather than on locating peaks, on avoiding dead-end streets rather than on computing the shortest distance, and on adaptability rather than optimum adjustment. This approach does not pretend to establish criteria for economic optimizing, but it offers a basis for water policy at successive stages of decision making.

DISCUSSION

ROBERT H. HAVEMAN: The significance of Professor Schmid's paper extends far beyond questions of water resource policy. Indeed, his exercise in concept-sorting has relevance for any and all social decisions in which questions of market failure or the distribution of income are at issue. The flavor of Schmid's analysis is distinctly unorthodox and, hence, refreshing. In treating each of the major market failure phenomena—intangibles, public goods, externalities, and indivisibilities—he operates out of no rigid set of presuppositions and attaches himself to no doctrinaire position, except, perhaps, one of shrewd pragmatism. In this discussion, I will attempt to filter out the essence of Schmid's position and, for better or for worse, to relate this position to that of standard or orthodox economic analysis.

While Schmid's analysis of intangibles is not unique, neither is it traditional. Whereas much standard analysis views these phenomena as immeasurable in dollar terms (e.g., Eckstein and McKean), Schmid, reflecting the tone of recent discussions on this problem (e.g., the recent Brookings Conference on Government Expenditures), urges that all such phenomena be at least conceptualized in the economist's framework of commodities, demand functions, and prices. Hence, investments which alter the quality of the environment, for example, require analysis in terms of the production of new or differentiated products each with a demand function and "price." In my view, the points which Schmid makes here are quite correct. Indeed, it is precisely this fixation on the nonanalytic quality concept which precludes rational discussion of reasonable social alternatives in the water pollution area. For example, were such questions approached as Schmid recommends, it is not at all unlikely that the reservation of some streams solely for waste disposal would form an obvious and essential element in a socially optimum pattern of resource use.

In his discussion of public goods, Schmid again shys away from the standard position which claims that such goods, because of their inherent characteristics, are nonmarketable and, hence, reserved for the public sector. Rather he suggests that a number of these commodities can be provided more effectively if institutions permitting group bargaining arrangements are devised and implemented. Relevant to Schmid's position here are recent contributions by Lind and Krutilla (*Water Resources Research*, Vol. 2, Nos. 2 and 4, 1966) proposing a compulsory flood loss insurance scheme as an institutional arrangement to both finance flood protection structures and regulate invasion of the protected flood plain. While the role of the public sector is not entirely eliminated in such a scheme, the enforcement of group bargaining subjects all flood plain beneficiaries to the exclusion principle and substantially eliminates flood plain protection from the public goods category.

The third set of nonmarket phenomena dealt with by Schmid are externalities. Here again the analysis significantly deviates from that of standard economic theory. Although disguised by his prose, the three standard categories of external effects—(1) technological externalities which are Pareto-relevant,

(2) technological externalities which are non-Pareto relevant, and (3) pecuniary externalities—are recognizable in Schmid's discussion. While standard externality theory holds that public intervention may be necessary to achieve a welfare maximum in the case of category 1 externalities (and then only if the Coase theorem is inoperative), it views both category 2 and 3 externalities as income redistribution and not real output effects. Accepting the proposition that the marginal utility of income is equal for all people, standard theory holds that no net welfare loss accrues from such externalities and, hence, they should be ignored in the social decision-making process. To Schmid's mind, however, all three categories are relevant to social decisions. In the case of category 1 externalities, Schmid demonstrates several instances in which the Pareto-better adjustment will not result without public intervention. He likewise claims decision-making relevance for both of the latter categories. In his view, it all hinges on the nature of property rights as designated by society; that is, the "welfare implication [of such externalities] depends on a [social] judgment of who is to count in a society." Once this judgment is made and rights are assigned, these externalities become as relevant to decision making as those in category 1.

As Schmid, I am sure, recognizes, such a position indeed opens Pandora's box. Surely, if the box were to be opened all of the way, the situation would be intolerable—for example, few income changes due to the market system adjustment process would be free from the demands for compensation. However, if the box could be cracked without jeopardizing the adjustment process in the private sector, the Schmid position would be of merit. As has been noted in recent applied welfare analyses in the water resource field, such an alternative is feasible by distinguishing those category 2 and 3 externalities imposed by public investment decisions from those inflicted by the market system adjustment process. Thus, for example, should the society decide that dollars of income to poor people are socially more valuable than dollars of income to rich people, the judgment on "who is to count" is formed and a basis for adjusting (weighting) efficiency dollars in public benefit-cost estimates achieved. Such a decision, while appropriate for inclusion in a multidimensional public investment criterion, need not be (and I would argue, should not be) applied to those externalities resulting from market system adjustment. That is to say, a deliberate policy of introducing adjustments for both non-Pareto relevant technological externalities and pecuniary externalities into the social investment criterion may well be in the public interest and worthy of implementation. Such a policy, however, need have no necessary implications for public interference to compensate for externalities imposed by the market system adjustment process. For Schmid's argument to be taken seriously (as I believe it should), this distinction must be made. Indeed, without it one has no rationale for dealing with the entire bundle of such externalities and the market system becomes stripped of its resource allocating function.

Much this same issue arises in the final section of Schmid's paper dealing with indivisibilities and other nonmarginal change phenomena. Because of such phenomena, Schmid correctly argues that existing prices no longer serve as guides to investment decision making and no unique social welfare opti-

mum emerges. Again he argues that a social judgment on property rights assignment must precede the formulation of optimal social policy. For Schmid, the task of the economist is to analyze the redistributive impact of alternative investments, the society's task is to determine how the impact should be distributed, and the task of the public sector is to undertake the activity. It all sounds so easy. Again the bothersome element in this position concerns the nature of reliable decision criteria in the absence of market signals. Even in the presence of such signals, the desirability of the outcome is often questionable. As Krutilla recently put it in evaluating the performance of the public sector in social investment decisions: "Perhaps not enough attention was paid the fact that while public intervention was *necessary* for remedying market failure, it need not be *sufficient* for improvement in efficiency. For intervention to be also a sufficient condition . . . appropriate criteria must be developed and, assuming in the final analysis that there is a feasible way to do so, applied with sufficient fidelity to insure that the objectives of public intervention in the interests of efficiency are reasonably approximated." (*Natural Resources J.*, Jan., 1966.)

Finally, I am struck by and am in hearty agreement with the sentiments present in all of the papers concerning the redirection of economists' efforts. It is indeed likely that supramarginal changes in effective resource policy require substantive work in the devising of institutions to implement the already developed conceptual framework. Regrettably, in neither the Schmid nor the Hirshleifer-Milliman paper does one find any concrete clues concerning the essential aspects of such a redirection. Without further defense, allow me to but suggest that major payoffs await the construction of institutional arrangements which tie beneficiary reimbursement to the appraisal and construction of public investments.

MASON GAFFNEY: I applaud and second the Hirshleifer-Milliman attack on municipal hydro-imperialism. Let cities develop and use resources close to home before jousting for remote supplies. I wonder if we would be having to maintain armed forces around the world today if we applied that philosophy in all our policies of resource development.

I applaud their touching on the role of water law in stimulating hydro-imperialism. The appropriative doctrine (first in time, first in right) creates the incentive pattern for allocating waters by a principle of comparative disadvantage. Comparative advantage would have a city use first those waters most convenient to it relative to their convenience for others. The appropriative doctrine motivates cities to grab first for waters most in jeopardy, which of course are those more convenient to rival cities.

I applaud their emphasis on the price elasticity of demand for water. My own observations support their postulate that water users can substitute water-conserving capital and management for water over a wide range and do so when water is scarce and/or high priced. I also observe with them that many engineer-planners are blocked and cannot perceive or even conceive of price elasticity, although they overstate the relative importance of this one factor.

I relish the irony of their kinship with Lange and Taylor (*Economic Theo-*

ry of Socialism), who also preached that public enterprises should use the price mechanism. Can it be that capitalist economists have more in common with socialist economists than with capitalist engineers and democratically elected "politicians"? Perhaps there lies another clue for world peace.

I applaud their emphasis on the problem of stochastic supply. Hardly any basic resource supply is so variable and unpredictable as water. Only the trough supply is "firm," and historical records are too brief to give us complete confidence in any alleged firm supply. Secular surface storage is prohibitively costly, so the need for safety factors is unusually high. None of that is news to water professionals, but it is news to many that price flexibility can substitute for a good deal of excess capacity in coping with the problem. Hirshleifer-Milliman are known for their support of peak-demand pricing to cope with demand variability. It is consistent that they offer us trough-supply pricing to cope with supply variability.

The alternative is for each city to grab so much natural supply that the minimum flow—say the hundred-year drought—covers the mean demand (or worse, the peak demand; worse yet, the future demand after a hundred years of rapid growth!). This would guarantee that most water would be surplus and wasted in all but the trough year. I wish they had given us some suggestions for how to work out better pooling arrangements to minimize the need for safety factors. The present pattern is one of almost pure vertical integration; i.e., each city controls its own supply and there is no pooling. Underground storage capacity is wasted even more; the law reserves it for the overlying landowners. There is "vertical integration" in the most literal sense!

On the negative side, I sorrow over the authors' penchant for overstatement; their willingness to generalize from two cases; their sometimes loaded language; their dark hints of unspecified "pressure" to explain decisions of which they disapprove; their use of "politicians and bureaucrats" as scapegoats for all the sins of municipal establishments; and their rhetorical posture of wide-eyed amazement at familiar policies. But I leave their enemies to capitalize on those weaknesses. I am concerned with supporting their basically sound case.

They might sharpen their analysis by using the common concept of rent. Superior waters, like other superior natural resources, bear rent. The problem before us is that those who control the superior waters do not always feel the rent as a price which would motivate economy. There is no explicit rent charged, because the state, the legal owner, gives away licenses to use free of charge; there is often no implicit opportunity price either, because water law creates no negotiable tenure instrument whereby cities or other municipalities (irrigation districts) with surplus waters can sell the surplus. On the contrary it blocks such alienation at every step. And so waters flow unvexed to the sea while cities go questing after submarginal sources.

What H-M say about "existing" supplies they really should, and perhaps intend to, say about rent-bearing supplies. Existing supplies may bear rent, but many are submarginal and should be abandoned in favor of unused superior waters. The behest to use "existing" supplies first is therefore not precisely on target, although well meant.

And why do cities develop submarginal waters that bear no present rent?

Because they expect the waters to bear future rent, and any future net value has some present value. Appropriative law, by granting future use as a reward for present use, motivates cities to establish their right of use thus as soon as the future rent is foreseen, long before there is present rent.

The fault does not lie with "bureaucrats" as such. Private water seekers respond in the same Pavlovian way as public ones to the stimuli in appropriative law. The fault lies in the basic process by which open-access resources—submarginal waters in this case—are reduced to tenure control. The prize of future rent-bearing use is established by present submarginal use.

The moral seems to be that the state should collect rent at the source for the use of its waters, including their nonuse as standby safety factors. The alternative is the "political" allocation and premature development which H-M deplore and the hogging of alleged safety factors which I deplore. They have elsewhere proposed that a free market in existing water rights be allowed to allocate waters, but without any assertion of the public equity. They pin their hope on implicit rent. I will not speak of the inadequacy of implicit rent to allocate natural resources: it is better than no market allocation at all. But the proposal does nothing to spare us the diseconomies inherent in the process by which new rights to tenure of water are continuously being created by appropriation from primordial submarginal chaos. So long as present appropriation secures claim to future rent, appropriation will be premature. State collection of water rents would prevent that abuse; it would also do much to prompt sale of surplus "existing" rights, and it would force a more sober appraisal by cities of their real need for safety factors.

The concept of rent also serves to underscore the parallel of water development and other resource development patterns. H-M rightly stress that water is not unique, but suggest that popular attitudes toward water are uniquely irrational. I suggest the pattern of underuse of rentable resources and premature development of submarginal ones is universal. It is clearly visible in oil, urban sprawl, agriculture, utility extension policy, area patterns of transportation rates, and all commodity cartel programs, for example. Water institutions have their idiosyncracies, but as to the underuse of rentable resources, *plus ça change, plus ça reste la meme chose.*

I wish they would patch up the seeming partial contradiction between their general support of price discrimination by monopoly water sellers, and their specific attacks on discrimination for agriculture and smaller suburbs. More generally, they need to open the whole question of area or zonal pricing as related to conveyance costs. The most common pattern of price discrimination is to charge equal rates for unequal costs of service, favoring out lands where the elasticity of demand for water is higher and aqueducts longer.

They should pat themselves on the back for their insistence, in 1960, on the use of high interest rates. Today's tight money underscores the folly of having frozen scarce capital in public projects yielding 2-3 percent when home buyers now are crying for money at 7 percent and more. Wouldn't it be nice to have those real capital assets back in disposable form today?

Finally, I think they misplace their emphasis when they conclude the solution is to throw the rascals out. What is the measure of malfeasance? Many

rascals have done well for their cities according to the rules of the game they are playing: appropriation law. We need new rules more than new rascals.

ARTHUR MAASS: Ciriacy-Wantrup argues, first, that for public investment analysis, political factors should be treated as genuine variables rather than simply as constraints on efficiency optimizing. I agree. But I am wary of the possible consequences, perhaps unintended, of C-W's response to his own argument; namely, that political variables should be introduced at a second level of system design and in a context quite different from that of quantitative optimizing.

Assume that the professionals in a federal bureau, on the initiative of the bureau chief and for the purpose of subsequent consideration by policy-makers in the executive and Congress, prepare a new program of public works for northern New England. Assume further that technological characteristics of the works and economic conditions in the nation and region are such that the program could provide benefits to the nation (efficiency, national income, or U.S. benefits) and/or to the region (regional income redistribution, or New England benefits). At the first stage of planning, the bureau men could collect data on the benefits of alternative projects and programs in a scheme like that shown in Table 1.

TABLE 1

SCHEME FOR COLLECTING DATA ON BENEFITS

	U. S. Benefits (national economic efficiency)	New England Benefits (regional income redistribution)
1. Project and program outputs	Willingness to pay of all users	Willingness to pay of users in New England *less* charges imposed on them
2. Underutilized resources used in construction and operation	Payments to underemployed resources *less* opportunity (real) costs of same	Payments to underemployed resources in region *less* any payments they will receive without program
3. Fully utilized resources attracted to New England for construction and operation	Ignore as benefits	Payments that remain in region
4. Incomes that result from increased spending by program beneficiaries (assuming full employment obtains *generally* in the U. S. economy)	Ignore	Incomes generated that remain in the region

Observe that although certain classes of benefits appear both in the U.S. and New England columns, this is not double counting. There are two objectives, and certain net changes in income within New England generate benefits with respect to both. At the same time, of course, we cannot simply add the two columns to obtain total benefits. U.S. and New England benefits are as

different as cotton and corn; they can be combined only with the aid of a trade-off ratio that expresses the policy-makers' preference for incomes in New England as against incomes elsewhere in the U.S.

It is at the second stage of planning that this political decision on the trade-off ratio can be made—in this case, perhaps, in legislation initiated by the President. In this respect my example is consistent with C-W's model. But you will note that my first stage involves a consideration of political (meaning in this case nonefficiency) variables; that my second stage involves quantitative optimizing, with appropriate weights; and that the context of decision making has significant continuity. The conclusions of this simple example have, I believe, general application and are not limited for relevance to certain specific problems.

Ciriacy-Wantrup argues, second, that we need to differentiate good from bad political decision processes; but that partisan mutual adjustment, the criterion used most frequently today to describe and evaluate these processes, does not satisfy the need. I agree. But here again I am wary of C-W's response to his own argument; namely, that analogies from biology are useful. As I read his admittedly speculative comments on this score, I am struck by the similarity between partisan mutual adjustment and survival value. In any case, I should think that a more direct approach would be to construct one or more models of decision making that are consistent with the basic requirements of a constitutional democratic state with representative government, and then study the extent to which model imperfections found in the real world (similar to market imperfections) can be tolerated or can be corrected.

Ciriacy-Wantrup argues, third, that decision rules for an integrated water system should account for public and private participation at several levels. I agree. The imprecise rules that we use today for this purpose are inadequate. Water resource plans have come to be judged by the extent to which they are comprehensive or multiple purpose—by how many of all possible purposes have been included. This is, of course, absurd. In terms of the community's needs and objectives it is more likely that a river should be developed for a few principal purposes than for all possible ones. In this case a decision rule, comprehensiveness, has been used to obscure rather than focus on the objectives of public action. Similarly, the planning process is judged frequently today by the extent to which it is coordinated; that is, the extent to which all conceivable interests at all levels of government have been given a voice in planning. But each of these interests has a different objective function, so that successful coordination in these terms introduces incredible complexity into planning. Here again a decision rule, coordination, has been used to obscure rather than focus on the objectives of public action.

Insofar as he develops it in his paper, I cannot see how the Ciriacy-Wantrup model can be used to write better decision rules than these, however. But this is a tentative conclusion; and I await anxiously the results of C-W's further research, for no one else has been able so far to write any better ones.

TRANSPORTATION AND PATTERNS OF URBAN DEVELOPMENT

AN AGGREGATIVE MODEL OF RESOURCE ALLOCATION IN A METROPOLITAN AREA*

By EDWIN S. MILLS

Johns Hopkins University

I. *Introduction*

The purpose of this paper is to put forth a simplified, aggregative model that will help to explain the sizes and structures of urban areas. The viewpoint taken is that the basic characteristics of cities are to be understood as market responses to opportunities for production and income. Properties of production functions are at the heart of the explanation of city size and structure in the model developed here, in much the way that properties of production functions are at the heart of modern neoclassical growth theory.

The general ideas that motivate the selection of the model developed below are commonplace in the voluminous recent literature on urban economics and geography. It has frequently been observed that the large size and rapid recent growth of urban areas are responses to income and employment opportunities provided there. It is but a small step from this observation to the assumption that the conditions of production differ in crucial respects as between urban and non-urban areas and as between urban areas of different size. Likewise, it is a common observation on the structure of cities that the nature and intensity of land use vary greatly from city to city and from one part of a city to another. Again, it is but a small step to recognize that a major element of factor substitution is involved in this phenomenon and to analyze models whose production functions will explain the observed factor substitution. Indeed, factor substitution is the most dramatic characteristic of urban structure. For example, the relative price of housing varies somewhat from one part of a city to another, but such variation is small compared with the variation in the relative prices of factors used to produce housing—principally land and structures. It is not unusual for land values to vary by a factor of from ten to one hundred within a distance of ten or twenty miles in a large metropolitan area. And the tremendous variation in capital-land ratios—from skyscrapers and high-rise apart-

ments downtown to single story factories and single family homes on two-acre lots in the suburbs—is the market's response to these dramatic variations in relative factor prices.

The model developed below is intended to shed light on these and other factors. To keep the mathematics within manageable proportions, it is necessary to make significant compromises with reality. In the work that follows, two major areas of compromise can be identified. First, the demand side has been slighted almost to the point of exclusion. This has been necessary in order to focus attention on what seem to me to be the crucial factors; namely, input substitution and technology. Second, the degree of aggregation is uncomfortably high. Even with these two areas of compromise, the model is quite cumbersome. Its solution is pragmatic and inelegant.

II. *A World without Cities*

It is clear that the existence, size, and structure of cities are closely related to transportation costs. The avoidance of transportation costs is not, however, a sufficient reason for the existence of cities. Indeed, it may help in focusing ideas to state explicitly a set of assumptions— each of which finds a respectable place in important economic models— which imply that there would be no cities.

Consider a general equilibrium model in which an arbitrary number of goods is produced either as inputs or for final consumption. The only nonproduced goods are land and labor, each of which is assumed to be homogeneous. Assume that each production function has constant returns to scale and that all input and output markets are competitive. Utility functions have the usual properties and have as arguments amounts of inputs supplied and products consumed. Under these circumstances, consumers would spread themselves over the land at a uniform density to avoid bidding up the price of land above that of land available elsewhere. Adjacent to each consumer would be all the industries necessary—directly or indirectly—to satisfy the demands of that customer. Constant returns assures us that production could take place at an arbitrarily small scale without loss of efficiency. In this way, all transportation costs could be avoided without any need to agglomerate economic activity.

III. *An Abstract Description of a City*

The two assumptions in the previous section most in conflict with reality are that land is homogeneous and that production functions all have constant returns to scale. Relaxation of either is sufficient to justify the existence of cities. Reasons for relaxing them and for the alternatives to them that are employed below are discussed in the next two paragraphs.

If some land is more productive than other land, it will pay to concentrate production on the better land, thus producing a city. The location of almost all U.S. cities can be understood in terms of land heterogeneity, most having been located near cheap water transportation. There are two ways to represent this heterogeneity in formal models. One is to assume that several variables related to land enter the production functions—natural resources, topography, climate, etc.—and that these variables are available in different amounts at different sites. Another is to assume just one land input, but to assume that different sites have associated with them different efficiency parameters in production functions. For a variety of reasons, the latter representation is chosen in this paper. With this convention, I would say that Baltimore's location results from the fact that some goods—especially transportation services —can be produced more efficiently there than further inland. The limited availability of desirable land will show up as decreasing returns as the amount of land used increases, forcing resort to less and less productive land. I will summarize this assumption by saying that efficiency parameters require locational indexes.

Location theorists have identified a variety of factors that lead to "agglomeration economies." The most important and best articulated of these factors is increasing returns to scale. This leads to agglomeration, not only of the activity in question, but also of other activities vertically related to it. Among other sources of agglomeration economies, most can probably be represented approximately as scale economies, at least in an aggregative model. Provided that the notion of scale economies is interpreted broadly, so as to include indivisibilities, it is undoubtedly important in determining city sizes. There are large numbers of specialized business and consumer services for which the per business or per capita demand is so small that a large city is needed to support even a few suppliers.

It is obvious that either locational effects on efficiency parameters or increasing returns will justify the existence of a city. Furthermore, conditions of production impose a finite limit on the efficient size of the city. Suppose we consider the possibility of doubling the population of a city by doubling the height of every building. If this were feasible and if twice as many people now traveled between each pair of points as before, then it would lead to just twice the demand for transportation as before. But if transportation requires land as an input, it must use more land after the doubling of population than before. Thus, some land previously used for buildings must now be used for transportation, thus requiring new buildings at the edge of the city. But the edge of the city has now moved out, and some people must make longer trips than before, requiring more transportation inputs. Thus, a doubling of the city's population requires more than doubling transportation inputs.

For a city of sufficient size, this "diseconomy" in transportation will more than balance any economy of size resulting from increasing returns in production. Another factor that entails the same result is the fact that, as the city's population grows, efficient production of goods requires the use of somewhat more land as well as of somewhat higher structures. At least this is true of any production function that has diminishing returns to factor proportions. Consequently, as a city grows, it moves out as well as up, and this entails diseconomy in transportation resources.

It was suggested above that the exhaustion of favorable land may show up as decreasing returns to scale in production. On the other hand, it was also stated that increasing returns in production is the most important agglomeration economy. It is thus important to formulate a model that is consistent with either increasing or decreasing returns to scale and to let the data tell us which assumption is appropriate.

IV. *The Model*

A. *Production Conditions.* The model developed here is an aggregate one. It assumes that only three activities take place in the urban area.

The first activity is the production of goods. The goods production function justifies the existence of the city. The city may be located where the efficiency parameter in the production function for goods is especially favorable. The production function may have increasing or decreasing returns. If there is no effect of location on the efficiency parameter, we must have increasing returns. Otherwise, there would be no city. If there are increasing returns, it is assumed that they are available only if goods production takes place in a contiguous area. If, instead, the city exists because of a site with a favorable efficiency parameter, then goods production will take place at this site. In either case, goods production will take place in a contiguous area and assumptions to be made below will imply that this area plays the role of the central business district (CBD). The production function is assumed to be Cobb-Douglas. Formally,

$$X_1 = A_1 L_1^{\alpha_1} N_1^{\beta_1} K_1^{\gamma_1} \qquad \alpha_1 + \beta_1 + \gamma_1 = H_1 \gtrless 1 \qquad (1)$$

where $X_1 =$ total output of goods, and L_1, N_1, K_1 are total inputs of land, labor, and capital in goods production. When written as in (1), the symbols refer to the amounts of inputs and outputs in the city. When reference is made to the value of a variable at a particular distance from the city center, the dependence on distance will be indicated explicitly. Thus, $X_1(u)du$ refers to the amount of goods produced in a ring of width du centered on a circle u miles from the city center. Then

$$X_1 = \int_{\text{city}} X_1(u)du$$

The second activity is intracity transportation. Assumptions will be introduced below that imply that the production of housing and other activities locate in "suburbs" around the CBD in a pattern determined by their bids for land. Transportation links the CBD with these suburbs. A great deal of factor substitution is possible in transportation. At one extreme, subways use little land but much capital. At another extreme, cars use much land but rather little capital. Probably the most realistic representation would be to assume that a choice must be made among a finite number of input-output coefficients relating inputs of land, labor, and capital to output of transportation. An efficient transportation system might then require the choice of a different set of coefficients in different parts of the city. However, the need for an integrated system places limits on this choice. Investigation of an optimal transportation system within the framework of the model developed here is a major goal of this study. The present paper, however, is restricted to studying the implications for city structure of choice of a particular set of coefficients. Thus, the coefficients are assumed to be exogenous in this paper. Actually, only one such coefficient is relevant for further analysis, as will be shown below. It is the ratio between land and transportation:

$$L_2(u) = bX_2(u) \tag{2}$$

$X_2(u)du$ is the number of passenger miles of transportation produced within a ring of width du u miles from the city center, and $L_2(u)du$ is the land input in transportation in this ring.

The third activity is designated "housing." The assumption is that all commodities whose production functions have nonconstant returns to scale and whose efficiency parameters are affected by location can be aggregated into the production function for goods (activity one). Competition will force the production of all other commodities to be located adjacent to customers in order to avoid transportation costs. It is assumed that the production of such goods and of housing can be aggregated into a single production function, designated "housing."

$$X_3(u) = A_3L_3(u)^{\alpha_3}N_3(u)^{\beta_3}K_3(u)^{\gamma_3} \quad \alpha_3 + \beta_3 + \gamma_3 = 1 \tag{3}$$

Once again, the u designates inputs and outputs within a narrow ring u miles from the city center. The assumption that all commodities can be dichotomized into the two groups designated as goods and housing is an approximation. In fact, there are degrees to which conditions of production require central location. Shopping centers display sufficient increasing returns to prohibit neighborhood location, but not enough to

require central location except in small towns. To introduce intermediate commodities of this sort would vastly complicate the model, since it would require each activity to be located not only with reference to distance from the center but also with reference to the distance from its neighbors.

B. *Market Conditions.* All factor markets are assumed competitive, so that each activity pays the same price for a given factor. Furthermore, the wage rate, w, and the rental rate on capital, r, are assumed to be exogenous. These are the appropriate assumptions if the city's size is to be endogenous. The city's population is determined by the number of workers it can bid for at the going wage rate. Likewise for the city's capital stock. The rental rate per acre of land u miles from the city center, $R(u)$, is endogenous.

Market conditions in the goods industry must be specified carefully. If there are increasing returns to scale, we cannot also have competitive product and factor markets. In this model it is assumed that the goods producer is a monopolist. The demand for X_1 is

$$X_1 = a_1 p_1^{-\lambda_1} \qquad \lambda_1 > 1 \qquad (4)$$

where λ_1 is the constant elasticity of demand. X_1 should be thought of as an "export" good. Alternatively, a_1 could be made a function of the city's population, although that possibility has not been incorporated into the subsequent analysis. A careful limiting operation in which λ_1 goes to infinity, but a_1^{1/λ_1} remains finite, would permit perfect competition to be included as a special case of (4).

It follows from the assumptions made in this model that the CBD will be circular. It can therefore be characterized by a single number, k_o, the distance from the city center to the boundary of the CBD. In order to increase land inputs for CBD uses, land must be bid away from suburban uses. Land rent at the boundary of the CBD, $R(k_o)$, therefore determines the use of land in the CBD. CBD land users take this rent as fixed, but its value will be determined by the model. Factor demands by industry 1 are thus determined by the following marginal productivity conditions:

$$\frac{\partial(p_1 X_1)}{\partial L_1} = R(k_o), \qquad \frac{\partial(p_1 X_1)}{\partial N_1} = w, \qquad \frac{\partial(p_1 X_1)}{\partial K_1} = r$$

Because of (4), these can be written as

$$\bar{\alpha}_1 \frac{p_1 X_1}{L_1} = R(k_o), \qquad \bar{\beta}_1 \frac{p_1 X_1}{N_1} = w, \qquad \bar{\gamma}_1 \frac{p_1 X_1}{K_1} = r \qquad (5)$$

where

$$\bar{\alpha}_1 = \epsilon_1 \alpha_1, \quad \bar{\beta}_1 = \epsilon_1 \beta_1, \quad \bar{\gamma}_1 = \epsilon_1 \gamma_1, \quad \text{and} \quad \epsilon_1 = 1 - 1/\lambda_1.$$

It is assumed that actual rent paid in the CBD just absorbs any monopoly profit. Thus,

$$R_1 = \frac{p_1 X_1 - w N_1 - r K_1}{L_1} = \frac{(1 - \bar{\beta}_1 - \bar{\gamma}_1)}{\bar{\alpha}_1} R(k_o) \tag{6}$$

the last equation following from (5). We must have $R_1 \geq R(k_o)$, otherwise industry 1 could not bid any land away from the surburbs. This inequality requires

$$H_1 \leq \frac{\lambda_1}{\lambda_1 - 1} \tag{7}$$

This inequality says that the greater the extent of increasing returns in goods production, the more inelastic must be goods demand in order to be able to pay the factors their marginal revenue products. It shows that the more competitive the goods market (the larger λ_1), the less the extent of increasing returns that is consistent with the model. In the limiting case, perfect competition requires constant or decreasing returns. (7) is assumed to hold in what follows.

Housing is assumed to be produced with competitive output—as well as input—markets. Thus

$$\alpha_3 \frac{p_3(u) X_3(u)}{L_3(u)} = R(u), \quad \beta_3 \frac{p_3(u) X_3(u)}{N_3(u)} = w, \quad \gamma_3 \frac{p_3(u) X_3(u)}{K_3(u)} = r \tag{8}$$

Using (3) and (8), we get the well-known expression for output price when markets are competitive and the production function is Cobb-Douglas.

$$\begin{aligned} p_3(u) &= [A_3 \alpha_3^{\alpha_3} \beta_3^{\beta_3} \gamma_3^{\gamma_3}]^{-1} R(u)^{\alpha_3} w^{\beta_3} r^{\gamma_3} \\ &= \bar{A}_3 R(u)^{\alpha_3}, \quad \bar{A}_3 = [A_3 \alpha_3^{\alpha_3} \beta_3^{\beta_3} \gamma_3^{\gamma_3}]^{-1} w^{\beta_3} r^{\gamma_3} \end{aligned} \tag{9}$$

It is assumed that housing consumption per worker is independent of u. Although this is not strictly correct, it is justified by the fact, stated above, that variations in the proportions in which land and capital are used to produce housing are much greater from one part of a city to another than are variations in the amount of housing consumed. We can express this assumption as

$$X_3(u) = N(u) x_3 \tag{10}$$

where $N(u)$ is the number of workers resident at a distance u from the center, and x_3 is the constant per worker housing demand.

It is assumed that a fraction ρ of the workers resident at each u is employed adjacent to their residences in the suburbs. It would be better to allow this proportion to be determined by the model, and presumably the conclusion would be that ρ would increase with u. Efforts to incorporate this possibility into the model have been unsuccessful. The assumption made amounts to the assumption that a fraction ρ of the workers resident at each u are employed in housing and transportation, and a fraction $(1-\rho)$ commute to the CBD. It is assumed that a transportation system adequate to handle these CBD commuters is also adequate for all other purposes. This is an accurate assumption for radial transportation and no other form appears in the model. With this assumption, the number of passenger miles of transportation needed at each $u \geq k_o$ is proportional to the number of workers who live beyond u and who commute to the CBD. By an appropriate choice of units, the factor of proportionality can be put equal to one:

$$X_2(u) = (1 - \rho) \int_u^{k_1} N(u')du' \qquad k_o \leq u \leq k_1 \qquad (11)$$

Here k_1 is the distance from city center to the outer edge of the suburbs. k_1 is endogenous. Likewise, the amount of transportation needed at a $u \leq k_o$ is proportional to the number of workers employed closer to the city center.

$$X_2(u) = \int_o^u N_1(u')du' \qquad 0 \leq u \leq k_o \qquad (12)$$

This ignores the commuting demand of transportation workers. This is legitimate if commuting is by car, since the commuters are then also the transportation workers. Otherwise it is an approximation.

It is assumed that the cost per passenger mile of transportation is proportional to $R(u)$:

$$p_2(u) = aR(u) \qquad (13)$$

This follows literally from (2) if it is assumed that land is the only transportation input. More realistically it is intended to reflect the fact that a major cost of intra-urban travel is the opportunity cost of time spent traveling and that travel is inevitably slower in denser, higher rent areas, even in an optimum transportation system. Although (13) is not necessarily the most realistic assumption that could be made, it greatly simplifies subsequent analysis.

A worker resident at u could decrease his transportation costs by moving in toward the city center. Equilibrium in the location of housing

requires that no such move be profitable. This will be so if the change in transportation cost from a short move is just offset by an opposite change in housing cost. This assumption can be expressed by the following equation:

$$p_2(u) + p_3'(u)x_3 = 0 \tag{14}$$

where the prime designates a derivative w.r.t. u. This crucial assumption appears in several models of urban location, but its implications appear not to have been analyzed.

The final assumption concerning market conditions is that urban users must be able to bid land away from some other uses, such as agriculture, at the edge of the urban area. Thus,

$$R(k_1) = R_A \tag{15}$$

where R_A is the opportunity cost of using land for urban purposes. R_A is exogenous, and (15) provides an "initial" condition for $R(u)$.

C. *Other Conditions*. Equilibrium requires that all land be used for some purpose. Within the CBD, land is used to produce goods and transportation, and we must have

$$L_1(u) + L_2(u) = 2\pi u \qquad 0 \le u \le k_o \tag{16}$$

In the suburbs, land is used to produce transportation and housing, and we must have

$$L_2(u) + L_3(u) = 2\pi u \qquad k_o \le u \le k_1 \tag{17}$$

(16) and (17) assume that there is no obstruction to a circular city. Topographical considerations—such as lakes, rivers, and harbors—may make a city of this shape impossible. If the obstruction is shaped like a pie slice, no fundamental alteration is necessary. If an obstruction takes up $(2\pi - \theta)$ radians at each u, then 2π can be replaced by θ in (16) and (17) and wherever 2π appears subsequently. Irregular obstructions cannot be handled within this model.

The relationship that completes the model says simply that all workers must live somewhere. This can be expressed as

$$N_i \equiv \int_0^{k_o} N_1(u)du = (1 - p) \int_{k_o}^{k_1} N(u)du \tag{18}$$

V. *Solution*

Despite the fact that the model presented in Section IV is drastically oversimplified in an economic sense, it is mathematically cumbersome. There does not seem to be any way of checking uniqueness or consistency by counting equations and unknowns, or any simple method of solution. Proceeding pragmatically and taking advantage of special

properties of the model, it is, however, possible to solve it. The endogenous variables are input and output quantities and prices in the three activities, the rent of land, and the distribution of residences—all expressed as functions of u. We should also be able to derive k_o and k_1. Of greatest interest are the expressions giving the rental value of land, the allocation of land among competing uses, and the density of population, each expressed as a function of distance from the center. Exogenous are the parameters of the three production functions, parameters of the demand function for goods, prices of labor and capital, the fraction of the labor force employed in the suburbs, the demand for housing per worker, and the rental value of land for agricultural purposes.

A. *CBD*. First, consider k_o and $R(k_o)$ to be fixed. Then from (5) we get the land-labor ratio in CBD goods production. Using (2), (5), (12), and (16), we get

$$L_1(u) + \lambda R(k_o) \int_0^u L_1(u')du' = 2\pi u \qquad \lambda = \frac{b\beta_1}{w\alpha_1}$$

Differentiating once, we get a first order differential equation in $L_1(u)$. Using the initial condition $L_1(0) = 0$, the solution is

$$L_1(u) = \frac{2\pi}{\lambda R(k_o)} (1 - e^{-\lambda R(k_o)u}) \tag{19}$$

This shows that the amount of land available for production increases at a decreasing rate as one moves out from the city center, despite the fact that the total amount of land available grows proportionately to u. The reason is that the land needed for transportation at u is proportionate to the integral of $N_1(u)$ up to u, and this grows much faster than u. Substituting (19) into (16), we have

$$L_2(u) = 2\pi \left[u - \frac{1}{\lambda R(k_o)} (1 - e^{-\lambda R(k_o)u}) \right] \tag{20}$$

If the city is sufficiently large, both k_o and $R(k_o)$ will be large. In that case, for large u, $L_2(u)$ is approximately $2\pi u$. This interesting result shows that, in a sufficiently large city, transportation will require nearly all the land near the edge of the CBD. But it cannot require more land than is available. That is, if CBD factor ratios are those dictated by competitive factor prices, the CBD will always be of a size such that there is enough land at the edge of the CBD to transport all those who work in the CBD.

This result also sheds an interesting light on CBD traffic congestion. Excessive congestion is not inherent in large city size. No matter how large the city there exists an allocation of CBD land that will avoid the

need for increases in passenger miles of transportation per acre of CBD land allocated to transportation. Congestion comes about because of the way cities grow. As a city grows (e.g., because of an increase in A_1 or a_1), $R(k_o)$ increases. As can be seen from (20), $L_2(u)$ is an increasing function of $R(k_o)$ for every u in the CBD. This is because an increase in $R(k_o)$ entails an increase in N_1/L_1 (and in K_1/L_1), which requires that a larger amount of CBD land be devoted to transporting the increased number of CBD workers. Congestion results because the adjustment of N_1/L_1 (and K_1/L_1) is relatively quick, whereas the transfer of CBD land from goods production to transportation is relatively slow. The former adjustment takes place mostly in the private sector, whereas the latter normally requires a transfer of land from the private to the public sector.

These and subsequent results can also be used to answer the following question, although the analysis has not been carried out. Suppose that CBD land is now allocated optimally, but that the city is expected to grow. Then three possibilities exist: (1) congestion will take place; (2) land will be transferred from goods production to transportation; (3) input-output coefficients in transportation must change (e.g., a switch from automotive to mass transit). What combination of the three is most economical? As it stands, the model considers only alternative (2), and it assumes that the city starts from scratch in that the cost of using CBD land for transportation is its rental value. For an existing city, the cost of transferring CBD land from goods production to transportation is its improved value, and this is much larger than its unimproved value.

So far we have considered only the input side of goods production. (19) tells us how much land will be available at each u in the CBD if factor proportions are those dictated by competitive factor prices. Taking account of the amount of X_1 that can be sold at the profit maximizing price, we get an expression for the demand for L_1. Making use of (1), (4) and (5), we get

$$L_1 = \bar{L}_1 R(k_o)^V$$

$$\bar{L}_1 = \left[A_1 a_1^{1/(\lambda_1-1)} \left(\frac{\beta_1}{\bar{\alpha}_1 w} \right)^{\beta_1} \left(\frac{\bar{\gamma}_1}{\bar{\alpha}_1 r} \right)^{\gamma_1} \bar{\alpha}_1 \frac{\lambda_1}{\lambda_1 - 1} \right]^{(\lambda_1-1)/(\lambda_1 - H_1(\lambda_1-1))} \qquad (21)$$

$$V = \frac{(\beta_1 + \gamma_1)(\lambda_1 - 1) - \lambda_1}{\lambda_1 - H_1(\lambda_1 - 1)}$$

It is easy to see that (7) implies $V < 0$, so that the higher are CBD land values, the less CBD land is demanded for goods production. It is also easy to check that an increase in A_1 or a_1 increases X_1 and L_1, as we should expect. We will return to (19) and (21) below where we will see how they can be used to determine k_o and $R(k_o)$.

B. *Suburbs.* Primary attention will be focused on finding the functions

$R(u)$ and $N(u)$. These are the most interesting variables from the theoretical and policy points of view. In addition, once these functions have been found all the input functions can easily be found using (2), (8), (9), (10) and (11).

$R(u)$ will be derived first. Substituting from (9) and (13) into (14), we get a differential equation in $R(u)$. Using the initial condition (15), the solution is

$$R(u) = [R_A{}^{-(1-\alpha_3)} - c(k_1 - u)]^{-1/(1-\alpha_3)} k_o \le u \le k_1$$
$$c = (1 - \alpha_3)a(\alpha_3 \overline{A}_3 x_3)^{-1} \tag{22}$$

It is sometimes asserted or speculated that land values fall off exponentially as one moves out from the city center. It is interesting to observe that exponential decline is a special case of (22) where $\alpha_3 = 1$. This is the special case where there is no factor substitution possible in housing; land is the only input. In general, (22) indicates a slower-than-exponential decline in land rents. Exponential decline means $R'(u)/R(u)$ is constant, whereas when $\alpha_3 < 1$, $R'(u)/R(u)$ is a decreasing function of u in (22). Another way to put this is to say that the possibility of economizing on land in housing prevents land values from rising as fast as they otherwise would as one moves in toward the city center.

Now turn to population (strictly, labor force) distribution, $N(u)$. Substituting from (2), (9), (10) and (11) in (17), we get

$$b(1 - \rho) \int_u^{k_1} N(u')du' + BR(u)^{-(1-\alpha_3)} N(u) = 2\pi u$$

Using the solution for $R(u)$ in (22), and differentiating, we get a differential equation in $N(u)$, whose solution is

$$N(u) = c_1 D^{-1}[BR_A{}^{-(1-\alpha_3)} - a(1 - \alpha_3)(k_1 - u)]^D - 2\pi D^{-1} k_o \le u \le k_1$$
$$B = \alpha_3 x_3 \overline{A}_3, \qquad D = b(1 - \rho) - a(1 - \alpha_3) \tag{23}$$

c_1 is an arbitrary constant of integration. Using the expression involving $L_3(u)$ in (8) and (9), (10), (15), and the initial condition $L_3(k_1) = 2\pi k_1$, c_1 can be evaluated as

$$c_1 = 2\pi k_1 B^{-(1+D)} DR_A{}^{(1-\alpha_3)(1+D)} + 2\pi B^{-D} R_A{}^{D(1-\alpha_3)}$$

It is to be noted in (23) that $N(u)$ is an increasing function of u if $D > 0$, and a decreasing function of u if $D < 0$. It is not possible to specify the sign of D a priori. This is as it should be. As u increases, three things happen: (1) the total amount of land increases proportionately to u; (2) the amount of land needed for transportation decreases; (3) population per residential acre decreases. (1) and (2) tend to increase $N(u)$, whereas (3) tends to decrease $N(u)$. The net effect depends on

the strengths of (1)–(3), and these are measured by the coefficients that make up D.

The behavior of population density is generally of greater interest than is $N(u)$. Net density[1] is $N(u)/L_3(u)$, whereas gross density is $N(u)/2\pi u$. Using the first expression in (8) and (9), (10) and (22) net density is given by

$$\frac{N(u)}{L_3(u)} = [BR_A^{-(1-\alpha_3)} - (1 - \alpha_3)a(k_1 - u)]^{-1} \qquad (24)$$

That is, the reciprocal of net density is linear in u. Colin Clark [1] has argued that population density falls off exponentially in all cities and at all times. Unfortunately, Clark does not make it clear whether he is using net or gross density and presents his evidence in a way that is difficult to evaluate. (No R^2's or significance tests are given, and there is no statement as to how the excluded CBD was determined.) Nevertheless, it is worth noting that no special case of (24) yields an exponential density function. Nor does any special case of this model yield an exponential gross density function.

The form of (24) makes empirical estimation and testing particularly easy. The reciprocal of net population density can be regressed on distance from city center. Furthermore, although the constant term in this regression (which depends on k_1) will vary from city to city, the coefficient of u should be the same for cities of different size. This provides an extremely simple partial test of the model.

Unfortunately, the effects of parameter changes on population density are not easy to ascertain within this model. Clark [1] asserts that an increase in transportation cost will increase population density near the center and decrease it further out. Within the model presented here, an increase in a represents an autonomous increase in transportation costs. The direct effect of an increase in a is as Clark states. But a change in a will also affect k_0 and k_1, and the net effect is difficult to ascertain. Among the questions one would like to answer, the most interesting would concern the effects of taxes and subsidies on the transportation system.

C. *Determination of k_0 and k_1.* All the solutions presented in the last two subsections, and solutions for other variables not presented, contain the two values k_0 and k_1 in addition to the autonomous parameters of the model. Equations to determine these values can be specified as follows.

k_0 must be such that the land available for goods production in the CBD equals the land that can be profitably employed in CBD goods

---

[1] Since land used for goods produced in the suburbs is included in $L_3(u)$, the measure of net density used here is not "as net" as the ratio of population to land used for housing.

production. The former is an integral of (19).

$$L_1 \equiv \int_o^{k_o} L_1(u)\,du = \frac{2\pi}{\lambda R(k_o)}\left[k_o - \frac{1}{\lambda R(k_o)}\left(1 - e^{-\lambda R(k_o)k_o}\right)\right] \quad (25)$$

The latter is given by (21). Equating the two gives

$$\overline{L}_1 R(k_o)^V = \frac{2\pi}{\lambda R(k_0)}\left[k_o - \frac{1}{\lambda R(k_0)}\left(1 - e^{-\lambda R(k_o)k_o}\right)\right] \quad (26)$$

Upon inserting the expression for $R(k_o)$ from (22), (26) becomes an equation involving only k_o and k_1 among the endogenous variables.

k_o and k_1 must also be such as to provide enough land for housing in the suburbs to house the workers who work there and those who work in the CBD. (18) ensures this. Upon substituting the solution for N_1 and the solution for $N(u)$, this too becomes an equation involving only k_o and k_1. Thus, (18) and (26) provide two equations for the two unknowns k_o and k_1. Although some progress has been made with approximations, the equations appear too complicated to learn much from them without resorting to numerical methods.

REFERENCES

1. Colin Clark, "Urban Population Densities," *J. of the Royal Statis. Soc.*, Series A, 1951, Vol. 114, pp. 490–96.
2. Richard Muth, "The Spatial Pattern of Residential Land Use in Cities" (unpublished manuscript).

THE LOCATION OF ECONOMIC ACTIVITY IN CITIES

By LEON MOSES, *Northwestern University*
and HAROLD F. WILLIAMSON, JR., *University of Illinois*

The decentralization or suburbanization of economic activity in major metropolitan areas, particularly those which grew to immense size during the nineteenth century, is a familiar phenomenon. The problems, such as increasing central city budgetary difficulties, associated with this process are also familiar. In this paper the authors hope to provide additional insight into the development of such large, core-dominated cities and into the factors which have and are affecting the location of economic activity within them.

The paper begins with a summary of a theoretical analysis developed to examine the structure of factor prices and costs within a core-dominated city. This analysis does not directly incorporate the influence of agglomerative economies, though their importance is acknowledged. The model highlights the effects of certain technological lags and transport cost relationships that the authors feel have not been sufficiently well understood. The model, moreover, assigns to the automobile a more modest role in the suburbanization of metropolitan activity than is found in most studies.

The remainder of the paper presents the results of two empirical analyses of the manufacturing sector in a particular core-dominated city—Chicago. (The analysis of location is made for this sector since manufacturing is the largest employer in many of these metropolitan areas and has had the largest central city decline in employment.) The first analysis presents evidence on an early stage of decentralization, roughly 1900 to 1920. The second summarizes some preliminary conclusions on the location of economic activity in a more recent period.

I. *Theoretical Analysis of the Core-Dominated City*

The city is divided into two areas: the core where all economic activity initially takes place and the satellite or residential zone surrounding it where a known number of households are located. The satellite area is assumed to be a transport plane and all households are assumed to be identical in terms of tastes—they have identical utility functions—and income earned at the core.[1] An equilibrium distribution of households requires that each is maximizing utility subject to constraints on income and time, the entire satellite ring is settled, and no

---

[1] The assumptions are similar to those used by other authors who have analyzed urban land rents. For example, see William Alonso's study [1] of land rent.

household can improve its level of satisfaction by changing location, meaning that all households are equally well-off.

It is the price of land which must vary so that these conditions are met since—as in the usual analysis of urban rent—incomes, transport costs, and the price of goods are fixed and known throughout the urban area. The result is a rent gradient—a function indicating the variation in the price of land as distance from the core varies. Since net income falls and the cost of goods rises as distance from the core increases, it can be shown that the rent gradient must decline with distance. Otherwise the equilibrium condition that utility is a locational constant would not be met.

Once the equilibrium distribution of households and the resulting rent gradient are established, a price gradient for labor—the wage rate gradient—can be determined. It measures the wage a firm has to pay at various locations in the satellite area to switch a given number of workers from core employment to employment at those sites, other things given, including the rent gradient.[2] The form of the wage rate gradient is not certain for all numbers of workers. It can, however, be shown to slope downward if a firm's employment were small relative to population in the vicinity of a potential site. The remaining factor price gradient, that for liquid capital, is invariant with respect to distance from the core. Capital is assumed to be perfectly mobile within an urban area so the interest rate is a locational constant.

The above analysis leads to the conclusion that factor costs tended to be lower in the satellite area. If this is the case, it is evident that other conditions must account for the growth of the core-dominated city. The agglomerative economies associated with proximity to competitive and complementary firms and to service industries are usually advanced to explain concentration of economic activity in the core. The authors acknowledge their importance but wish to focus attention on the structure of transport costs in the nineteenth century, and the influence this structure had on the form of large cities. It is a thesis of this paper that in the nineteenth century the cost of moving goods within cities was: (1) high relative to the cost of moving people within cities; and (2) high relative to the cost of moving goods between cities. Location in the satellite area involved moving away from the central goods handling facilities of the city. In general, the increased

---

[2] For a detailed development of the analysis concerning this gradient, see Leon N. Moses [11]. The assumption that the rent gradient is held constant means that the impact of the firm's relocation to an alternative site is not considered. This impact includes the changes in work trip lengths for the affected households and the development of a second rent gradient centered at the new employment site. This is not considered here since the analysis is sufficient without it to provide insights into the location of urban plants and the factors affecting their movement.

costs of transporting inputs from and outputs to the core (for ship-
ment to other areas) outweighed the savings from the lower factor
prices associated with location in the satellite area.[3]

This in fact tended to be the situation during the period when the
core-dominated cities grew. Firms received from and shipped to other
regions a significant proportion of their inputs and outputs. The econo-
mies of scale in rail transport—the main mode for interregional goods
movement—were such that the receiving and sending of such ship-
ments were concentrated at one or at most a few large, centrally lo-
cated freight terminals. Within cities, the movement of people was rel-
atively efficient, being carried out by modes which were closely related to
the railroad: trolleys and street railways. Intracity movement of
goods, however, took place by an inefficient mode, the horse and
wagon. The cost of moving goods was, therefore, high relative to the cost
of moving people. This relative cost relationship played a crucial role
in the emergence of the core-dominated city. The lower transport costs
associated with location in the core exceeded the reduction in cost pos-
sible from lower wages and rents at sites in the satellite area. A prereq-
uisite for decentralization was the breaking of the transport tie to the
core.

Only after technological changes occurred in transportation was the
attraction of a non-core location strongly felt. The major change was
the introduction of the truck which reduced the cost of moving goods
within cities. Its effect on the spatial structure of cities can, roughly,
be divided into two phases. During the first, the motor truck was intro-
duced and became the dominant form of intra-urban carriage, but in-
terurban carriage was still done by railroads. In this period—the first
two decades of this century—firms could leave the core but were still
tied to it for shipments to and from other regions. This tie was weak-
ened during the second phase when improvements in the truck and in
the interregional highway system meant this mode could be used for
long-distance transport. The full impact of this change was probably not
felt until the revival of a strong peacetime economy after World War II.
The attractiveness of the satellite area in this period was increased by
the automobile which allowed firms to draw labor from a broad area.[4]

In summary, the theoretical analysis emphasized the effect of
changes in transportation cost structure. It implied that decentraliza-

[3] For a study of freight costs and their impact upon the New York Metropolitan Re-
gion, see Benjamin Chinitz [6, Part II, particularly Chap. 6].
[4] The attractiveness of suburban locations was also enhanced by other technological
changes during this period; for a discussion of such factors, see J. R. Meyer, J. Kain, and
M. Wohl [10, pp. 10-18]. Improvements in data processing and communications, for ex-
ample, meant that firms could remain "near" other firms though they had moved many
miles away.

tion should have begun, roughly, during the first two decades of this century when the motor truck was first introduced, though this period is not usually thought to be one of suburbanization. The first part of the empirical analysis examines this period. The second part provides a more extensive examination of the spatial rearrangement of firms during the post-World War II period—the period during which the impact of interregional transportation was felt.

II. *Empirical Analysis: The Early Period*

There is fragmentary evidence in support of the proposition that the introduction of the truck in the first two decades of the present century did reduce the cost of moving goods within cities relative to the cost of moving people. The cost of moving people—to judge from the fare —was fairly constant during this period. The passenger fare was almost universally five cents at the beginning of the century and remained in force until approximately World War I.[5] At the same time, motor carriage began to replace the horse and wagon in urban goods transport. Evidence of the relative efficiency of the two modes is seen in the rapid rise in truck registrations (800 in 1910 in Chicago to 23,000 in 1920) and decline in the number of horse-drawn vehicle registrations (58,000 to 31,000 during the same period).[6]

Reliable data on cost differences between motor truck and wagon are practically nonexistent for urban areas but Department of Agriculture surveys of farm use give some idea of relative costs. A survey in 1918 found that horse and wagon costs were $.33 per ton-mile. This was a bit higher than similar surveys had found it to be near the turn of the century. The ton-mile cost for truck was $.15. Another survey in 1920 found average hours per ton-mile were one-third to one-half as great for motor truck as for horse and wagon. The truck thus appears to have had a time as well as a direct money cost advantage.[7] These findings were, however, for hauls longer than those common in urban transportation during the period. They are, therefore, somewhat biased against the horse and wagon since its disadvantages increase as distance increases. Even with this qualification, the data do give rough evidence that the introduction of the motor truck reduced the cost of moving goods during a period when the cost of moving people was relatively constant.

Evidence of decentralization during this period of changing transportation costs is also limited. It is based on a survey made by the au-

---

[5] See American Transit Association [2, Chap. I].

[6] Homer Hoyt [8, p. 485].

[7] The money results from the 1918 survey are from George W. Grupp [7, pp. 38-39]; while the time results from the 1920 survey appear in H. R. Tolley and L. M. Church [12, p. 11].

thors of 955 firms in four industry groups.[8] Information on location was obtained for 1908 and 1920.[9] There were 473 firms which could be identified at both dates. Two hundred eighty-five of these firms had shifted location during the period.

The behavior of these firms is in agreement with the conclusions reached in the theoretical analysis. While firms moved a short distance on average, the direction of movement was away from the core. The average distance from the core for firms which moved was 59 percent greater in 1920 than it had been in 1908, having increased from 0.92 miles to 1.46 miles. In addition, the average origin (1908) distance for firms which moved was less than the distance for nonmoving firms. This agrees with the implication that the introduction of the motor truck had a greater impact on core area firms than those already located somewhat outside the core.

III. *Empirical Analysis: The Recent Period*

For the years 1950 through mid-1964, it was possible to obtain rather detailed information for 2,000 firms in the Chicago area which had either relocated or expanded capacity at a site different from their existing one.[10] Information on the size, location, and industry group of each firm was gathered from the files of the Chicago Association of Commerce and Industry and the Illinois State Employment Service. Destination for each movement was plotted by Chicago Area Transportation Study (CATS) zone and also by postal zone within Chicago and suburbs outside the city. Data on characteristics of the metropolitan area such as population and land use were obtained from CATS surveys and maps. With this information, it was possible to analyze the pattern of location and movement. This analysis is divided into three parts: the distribution of origins, the direction and distance of movement, and the pattern of destinations.

Distribution of Origins. Preliminary investigation revealed that the distribution of origins fell off sharply as distance from the core of the city increased. One interpretation of this result is that inner portions of the city have become undesirable as locations, so that firms located

[8] The industry groups were printing and publishing, chemicals, nonelectrical machinery, and electrical machinery. They were chosen because of their importance to the city's economy and because firms in these industries could be most easily identified in the industrial directories used for gathering the sample.

[9] Though it was hoped to cover the first two decades of this century, the study was begun in 1908 because this was the first year for which a directory was available which provided an adequate industrial identification for the firms. The directories used were the *Membership Directory* (1908) and the *Blue Book of Chicago Commerce* (1920), both issued by the Chicago Association of Commerce.

[10] The metropolitan area was defined to be the study area of the Chicago Area Transportation Study (CATS) and the geographic divisions used were those developed by CATS. For information on the geographic, land use and population characteristics of the study area, see CATS [3] [4] [5].

there have a higher propensity to move. This conclusion is somewhat misleading in that it ignores the fact that zones near the core also are likely to have larger numbers of firms in them. The critical issue, therefore, is not whether number of origins declines with distance but whether the percentage of origins behaves in this manner.

To see whether number of firms or distance was more highly correlated with number of origins, regressions were run with origins in each zone as the dependent variable and these two factors as the independent variables.[11] The geographic divisions used were postal zones within the city of Chicago and suburbs outside it. When number of origins was regressed on each variable alone, number of firms provided a much higher explanatory power (coefficient of determination between .51 and .90) than did distance (coefficient of determination between .19 and .43). When the regression was run for both independent variables, number of firms had the higher explanatory power. The additional explanation provided by the distance variable was not significant in over half the cases. Thus, number of origins is highly correlated with number of firms in a zone. The coefficient for number of firms was from ten to twenty times as large as its standard error; i.e., the t ratio was between ten and twenty. Economic dispersal has not therefore occurred because firms that are closer to the core have a higher propensity to move. Instead, this propensity is fairly constant over the entire metropolitan area. Since this percentage of firms being "set loose" is relatively constant, the shifting pattern of industrial location must result from the spatial pattern of destinations—the percentage of firms which "set down" in each zone. Before turning to the analysis of this pattern of destinations, it is necessary to determine whether these firms are truly set loose from their origin site.

Distance of Movement. The preceding empirical work emphasized net change in distance from the core; i.e., distance from the core of a firm's destination minus its origin distance. Net change is a surrogate for the variation in land and labor costs and for the expense of maintaining linkages with the core which arose due to the move. A different measure of the connection between origin and destination is the distance actually moved by the firm. It reflects the costs associated with the linkage between these two sites. As distance moved increases, established ties with suppliers of raw materials and services, labor supply, and customers may be attenuated. Costs may then have to be incurred to establish new ties. If so, firms would not be set loose when they decided to move. Instead, there would be factors unrelated to the

[11] Though the results were summarized without this distinction, there were actually four groups of firms examined. The division was by type of move—relocation or expansion at new—and by period—1950-59 and 1960-64. The latter division was made because the data for each period were gathered at different times by different groups though from the same sources. In general, the results are similar for all four groups.

attributes of potential destination areas which influence whether loca-
tion there is optimal. The distance which a potential destination is
from the firm's origin appears to be a good proxy for these factors.

The distribution of firms falls off sharply when arranged by distance
moved. The median distance moved for the four categories of firms de-
scribed in footnote eleven ranged from four to six miles. The relation-
ship resembled that of a gravity model in which the number of inter-
connections between zones diminishes as the distance between them
increases. A regression fitted to the data for all firms in the sample
yielded a significant negative relationship between number of firms
and distance moved, particularly when the latter was expressed loga-
rithmically. The coefficient of determination was at least .50 in all
cases and between .81 and .95 for the logarithmic form. Further in-
sight into distance moved was obtained by examining various charac-
teristics of the firms.

Size of the firm influenced distance moved.[12] Smaller firms tended to
move shorter distances than larger ones. This would seem to support
the conclusion that distance moved is generally short because it
reflects the cost of moving. Larger firms can move longer distances,
perhaps because they are more independent of suppliers or buyers at a
particular location.[13] Size should therefore be taken into account when
analyzing location patterns, since smaller firms may not be free to
choose among all possible destination areas. Instead, they will be lim-
ited to those within a relatively short distance of their origin location.

Two conditions which did not seem to affect the pattern of distance
moved were the origin distance of the firm and its industrial category.
A regression with distance moved as a function of origin distance was
run and the coefficient of determination was never as high as 6 percent.
The search for an industry effect had to be conducted in terms of the
two-digit Standard Industrial Classification in order to have enough
observations. At this level of aggregation it was found that the dis-
tribution of firms by distance moved for each industry was never
significantly different (at the .05 level) from the distribution of all
firms.

The Analysis of Destinations. Decentralization is due largely to the
pattern of destinations of firms that shift location.[14] The purposes of
this section are to explain the method by which the authors are exam-

[12] Size was measured by the cost of land and construction associated with relocation or
expansion at a new site.

[13] The regression of number of firms on distance moved provided a good fit for the small
size categories but had a reduced and often insignificant coefficient of determination for
large firms. In other words, the distribution of number of firms did not decrease sharply
as distance moved increased for the latter group.

[14] It also reflects locational patterns of firms that go out of business and of firms that
are new to the area. Though a limited set of data is available for the latter group, it is
not considered in this paper.

ining the pattern of destinations, to present some preliminary results, and to indicate the direction of additional empirical efforts.

The basic tool employed in the analysis of destinations was the following explanatory equation:

$$D = a + b_1L + b_2W + b_3T + b_4H + b_5V + b_6M + b_7C + u$$

where a is the constant term and u the error term. D, the dependent variable, is the density or number of destinations per unit area for the 582 CATS zones used as observations.

The independent variables are measures of the various factors which affect the location decision. The first two are proxies for the rent and wage gradients, since it is nearly impossible to measure either of these.[15] In place of the rent gradient, the distance of the zone from the core, L, is used since this is inversely correlated with the rent gradient.[16] A surrogate for the wage gradient is the availability of labor for firms locating in a given zone. An initial measure of this is W, population density in the given zone.[17]

The percentage of land in a zone used for transportation other than highway, T, was used to measure access to such modes as rail and water. For highway service, the stress was upon availability of limited access freeways or tollways. Accessibility was measured by H, a dummy variable which was one if such a highway was in the given zone or one adjacent to it, and zero if this were not the case. The availability of land was measured by V, the percentage of vacant land zoned for manufacturing and commercial purposes, and M, the percentage of land in manufacturing use. The final variable, C, is a dummy

[15] Assessed valuations of land provide some information on the price of land but the problem of forming an index of price to account for variations in these prices and type of land seems insolvable. Wage rate figures, on the other hand, are rarely broken down for sufficiently small areas within the metropolitan area to be useful here, and the problem of forming an index if they were would be equally difficult.

[16] A variant of distance from the core was used, for example, by John Kain [9] to approximate the relative price of residential space in an urban area. Distance from the core is actually a proxy for two, inversely correlated independent variables. The first is the rent gradient as theoretically derived from the household sector equilibrium. This gradient is negatively correlated with both distance and destinations. The second is the cost gradient reflecting the difficulty of contact with the core and the services offered there. This cost gradient is positively correlated with distance but negatively correlated with destinations. It is necessary to consider this dual interpretation of the distance variable when discussing the meaning of its coefficient.

[17] It is intended to improve this variable by using, instead, an index of availability of labor to any zone j. This index will be the sum of the population in each zone i weighted by an inverse function of the distance or travel time of this population from zone j. Limited support for this proxy can be found in a Chicago labor market study being conducted by Professor Albert Rees at the University of Chicago. When distance to work was included as an explanatory variable for the wage rate, it was found to have a positive effect. The farther workers had to travel, the higher was their wage. Thus, a readily available labor force might indicate a lower wage rate. Even if wage rates are relatively constant in the urban area, availability may be an attracting factor since it can result in a reduction in the costs of finding and hiring new workers.

variable which is zero if more than half the zone is within the city of Chicago and one if it is not. This is the major political division in the study area. It was introduced to catch the effects of such things as differences in zoning policy, property tax rates, etc., between the central city and the rest of the area.

This equation was run for the firms which moved during the years 1950-59 because data for the explanatory variables were collected during that period. The results are presented in Table 1. The first two rows indicate the values when the equation was run for expansions at new locations and relocations, respectively, for the entire area (582 zones). As indicated by the coefficient of determination (R^2), this equation accounted for between one-fifth (.2152) and one-fourth (.2471) of the variation in the dependent variable. The only significant variables were distance, L, and percentage of manufacturing land, M.[18] Before discussing them, it should be noted that all but one of the remaining coefficients have the expected sign. Other things equal, an increase in population or transportation land, the existence of a limited access highway near the zone or a change from the city to the suburbs will be associated with an increase in destinations. Only vacant land has an impact (negative) which is different from that expected. This may be because the inaccuracy of zoning definitions makes this variable a poor indicator of the availability of destination sites. The significant, positive coefficient for manufacturing land may indicate that this variable, instead, provides a measure of such availability. The significant, negative sign for the distance coefficient seems to indicate that it is a measure of the attraction of the core rather than a proxy for the rent gradient.

Though these explanations are logical, another more likely one for the signs of the latter two coefficients is suggested by the results of the distance moved analysis. Since most firms move short distances and most origins are near the core, distance and manufacturing land may merely be measuring the proximity of a zone to firms which are moving. Thus, increases in distance reduce the number of destinations since the zone is farther from firms which are moving. Similarly, as the percentage of manufacturing land rises, destinations rise, since more firms are located in that square. The ability of the equation to measure the influence of all the locational factors examined may be improved by introducing a variable for proximity. Two approaches suggest themselves. The first is to divide firms by size group and examine the equa-

[18] Significance is indicated by ** (.01) or * (.05). Although the necessary assumptions for such tests of significance may not be met, it is assumed that the ratio of the coefficient to its standard error (the standard error appears in parentheses below the value of the coefficient)—the t variable—provides a roughly accurate measure of the variable's significance.

TABLE 1

REGRESSION RESULTS FOR EQUATION ONE

Type of Analysis	(1) R²	(2) Intercept a	Values of the Coefficients for the Variables†						
			(3) L	(4) W	(5) T	(6) H	(7) V	(8) M	(9) C
1. Expansion at new, total area......	.2471	1.778	−.0956 (.0174)**	.0177 (.0116)	.7671 (.9306)	.2260 (.1607)	−.6144 (.8615)	−6.8012 (1.2145)**	.1490 (.2458)
2. Relocation total area......	.2152	1.501	−.0829 (.0153)**	.0148 (.0102)	.3412 (.8187)	.2610 (.1413)	−.4791 (.7579)	4.5580 (1.0684)**	.1097 (.2612)
3. Expansion at new, north......	.6600	.481	−.0354 (.0259)	.0362 (.0190)	−4.3679 (2.7080)	.3317 (.2662)	−4.6234 (4.3808)	42.9550 (3.5370)**	.2442 (.4112)
4. Expansion at new, west......	.4346	2.091	−.0349 (.0292)	−.0066 (.0241)	1.0869 (1.1848)	.1451 (.2581)	−.4858 (2.2110)	12.1229 (2.3806)**	−1.4064 (.5055)**
5. Expansion at new, south......	.2801	1.303	−.0487 (.0168)**	−.0064 (.0103)	3.1634 (.8577)**	−.5261 (.1377)**	−.0313 (.5625)	1.3638 (.8369)	−.0789 (.1994)
6. Relocation north...	.5536	.121	−.0162 (.0220)	.0434 (.0161)	−1.7338 (2.2997)	.1845 (.2261)	2.4800 (3.7202)	24.9222 (3.0032)**	.1639 (.3492)
7. Relocation west....	.4662	1.482	−.0292 (.0220)	−.0052 (.0181)	.8782 (.8901)	.1422 (.1939)	−.5638 (1.6613)	10.6777 (1.7885)**	−.9129 (.3797)**
8. Relocation south...	.2521	.830	−.0358 (.0132)**	.0028 (.0081)	2.2491 (.6727)**	−.2696 (.1080)*	.0988 (.4412)	.9438 (.6564)	−.0596 (.1564)

† Explanation of variables: L, distance in miles; W, population as thousands per square mile; T, percentage of land in transportation; H, dummy variable: one if highway, zero if not; V, percentage of vacant land; M, percentage of manufacturing land; C, dummy variable: zero if zone within city, one if not.

The number in parentheses below the value of the coefficient is the standard error. If the coefficient is significantly different from zero at the .05 level of significance, it is indicated by * while ** indicates a .01 level of significance.

tion for each separately. A second is to introduce a measure of proximity of each zone to firms. This measure could be similar to the one suggested above for availibility of labor.

The importance of this and other improvements in the analysis can be seen from the results in the rest of Table 1. The above equation was fitted to data for each of three sectors into which the city was divided.[19] This division was made because examination of the data indicated noticeable differences in these sectors, particularly with respect to manufacturing land. Though this variable had a significantly positive coefficient in all but one of the regressions, the sector with the highest percentage of such land, the south, had the lowest density of destinations. The reverse was true for the north.[20] The result of the division into sectors is a noticeable improvement in the explanatory power of the regression, particularly for the north and west sectors.[21] This improvement in the statistical results is not conceptually significant in itself. It indicates that there are critical variables that have not been included. It is hoped that with their identification it will be possible to obtain results for the entire study area that are as good as those that have been obtained for the somewhat arbitrary geographic subdivisions of it. One of these variables is suggested by the results of the sector analysis. As has been noted, zones in the southern sectors tend to have a higher percentage of manufacturing land than zones in the remainder of the study area. Manufacturing land always has a significant, positive effect on destinations yet the southern sectors attract the fewest number of firms. The explanation may be that zones in these sectors also tend to have a higher proportion of nonwhite population. A variable that reflects population composition of zones will, therefore, be included in future empirical work.

[19] The sectors were formed by dividing the metropolitan area outside the core (CATS ring-sectors 01 and 11) into the north sector (156 zones in CATS sectors 1 and 2); the west sector (136 zones in CATS sectors 3 and 4) and south sector (270 zones in CATS sectors 5, 6, and 7)

[20] The number of destinations in both groups per square mile in the north, west and south sectors, respectively, was 1.21, 1.00, and 0.54, while the percentage of land in manufacturing was 1.9, 3.1, and 4.4.

[21] When the regression is run for each of the seven CATS sectors, the explanatory power is further improved, especially in the southern sectors. Except for one case, the coefficient of determination is over 0.40 and ranges as high as 0.74.

REFERENCES

1. William Alonso, *Location and Land Use: Toward a General Theory of Land Rent* (Harvard Univ. Press, 1964).
2. American Transit Association, Committee on Fare Structures, *Fare Structures in the Transit Industry* (New York, 1933).
3. Chicago Area Transportation Study, *Geographic Reference System* (20,000) (Chicago, Aug., 1963).
4. ———, *Home Interview Population Manual* (31,210) (Chicago, June, 1959).
5. ———, *Land Use Survey Manual* (15,210) (Chicago, 1956).
6. Benjamin Chinitz, *Freight and the Metropolis* (Harvard Univ. Press, 1960).

7. George W. Grupp, *Economics of Motor Transportation* (Appleton & Co., 1924).
8. Homer Hoyt, *One Hundred Years of Land Values in Chicago* (Univ. of Chicago Press, 1933)
9. J. F. Kain, *A Multiple Equation Model of Household Locational and Tripmaking Behavior* (RAND Corp., Memorandum RM-3086-FF, Apr., 1962).
10. J. R. Meyer, J. F. Kain, and M. Wohl, *The Urban Transportation Problem* (Harvard Univ. Press, 1965)
11. Leon N. Moses, "Toward a Theory of Intra-Urban Wage Differentials and Their Influence on Travel Patterns," *Papers of the Reg. Sci. Asso.*, 1962.
12. H. R. Tolley and L. M. Church, "Experience of Eastern Farmers with Motor Trucks," U.S. Dept. of Agric. *Bul. No. 910* (Washington, D.C., 1920).

POSTWAR METROPOLITAN DEVELOPMENT: HOUSING PREFERENCES AND AUTO OWNERSHIP*

By John F. Kain

Harvard University

U.S. metropolitan areas have been literally transformed in the two decades since World War II. Extensive geographic growth of metropolitan regions, employment and population declines in central areas, and low density development, particularly residential, are perhaps the most notable dimensions of this transformation. While metropolitan area population grew by 26 percent between 1950 and 1960, mean central city densities declined from 7,800 to 5,800 persons per square mile.[1] During the same period auto ownership increased from a level of 0.69 per household in 1945, to 0.92 in 1950, and to 1.16 in 1960.[2] This coincidence of rapid increases in automobile ownership and transformations in metropolitan structure have caused many observers to conclude that the growth in automobile ownership is the primary cause of postwar changes in metropolitan structure.

There is one notable and striking exception to this view of the relationship between automobile ownership and metropolitan development. Elaborate urban transportation studies have been carried out in over 200 U.S. metropolitan areas since World War II. In these studies, automobile ownership is used to explain and predict levels of tripmaking, choice of travel mode, and other aspects of urban travel. But auto ownership is almost always assumed to depend on net residential density or just the opposite assumption of causality from that noted

[1] U.S. Bureau of the Census, *U.S. Census of Population: 1960, Number of Inhabitants, U.S. Summary. Final Report* PC (1)-1A (U.S. Government Printing Office, 1961), p. 1-40.

[2] Automobile Manufacturers Association, *Automobile Facts and Figures*, 1964 Edition, p. 18; United States Bureau of the Census, *Statistical Abstract of the United States: 1964* (85 ed., U.S. Government Printing Office, 1964).

above. Projections of net residential density in turn are made without consideration of automobile ownership.[3]

This paper addresses itself to the reconciliation of these conflicting views about the causal relationships between auto ownership and net residential density. Econometric models are estimated based on the hypotheses that: (1) residential density depends on auto ownership per household, income, and preferences for residential space with unidirectional causality from auto ownership to density; (2) auto ownership per household depends on residential density, income, and household transportation requirements with unidirectional causality from residential density to auto ownership; and (3) automobile ownership and residential density are jointly and simultaneously determined.

Variables

Empirical evaluation of these alternative hypotheses is based primarily on data for 54 cities and towns in the Boston Metropolitan region for 1950 and 1960. The variables used are listed below.[4] All except the auto ownership ratio, which refers to 1963 only, are defined for both 1950 and 1960. Both 1949 and 1959 median family income are defined in terms of constant 1959 dollars.

1. A = Automobile ownership per hundred household (Total Polk automobile registrations times 100/Total households).

2. D = Net dwelling unit density (Total dwelling units/Estimated number of acres in residential use).

3. S = Percent single family (Number of single family dwelling units times 100/Total dwelling units).

---

[3] For a survey and discussion of automobile ownership in these studies see: John F. Kain, "Urban Travel Behavior," in Leo F. Schnore and Henry Fagin, eds., *Urban Research and Policy Planning: The First Urban Affairs Annual Review* (forthcoming); also available as *Discussion Paper Number 2*, Program on Regional and Urban Economics, Harvard Univ., Sept., 1966.

[4] Housing structure, income, and population data for 1950 were obtained from: U.S. Bureau of the Census, *U.S. Census of Housing: 1950*, Vol. I, *General Characteristics*, Chapter 21, Massachusetts (U.S. Government Printing Office, 1952); *idem, U.S. Census of Population: 1950*, Vol. III, *Census Tract Statistics*, Chap. 6, Boston (U.S. Government Printing Office, 1952); *idem, U.S. Census of Population: 1950*, Vol. II, *Characteristics of the Population*, Part 21, Massachusetts (U.S. Government Printing Office, 1952). Housing structure, income and population data for 1960 were obtained from: U.S. Bureau of the Census. *U.S. Census of Housing: 1960*, Vol. I, *States and Small Areas*, Massachusetts (U.S. Government Printing Office, 1961); *idem, U.S. Censuses of Population and Housing: 1960, Census Tracts*, Boston (U.S. Government Printing Office, 1962); *idem, U.S. Census of Population: 1960, General Social and Economic Characteristics*, Massachusetts (U.S. Government Printing Office, 1961). Automobile registration data for 1950 and 1960 were obtained from motor statistics provided by R. L. Polk Company, Cincinnati, Ohio. Land use data for 1952 and 1963 and employment data for 1950, 1960, and 1963 were supplied by the Metropolitan Area Planning Council, Boston, Massachusetts. Dwelling units, cars owned, total residents, and employed residents data for 1963 were obtained from Wilbur Smith and Associates, *Comprehensive Traffic and Transportation Inventory*, (New Haven: Wilbur Smith and Associates, Sept., 1965).

4. $Y =$ Median family income in hundreds of constant 1959 dollars.

5. $F =$ Mean family size (Total population in households/Total households).

6. $L =$ Labor force participation rate (Total employed residents/Total households).

7. $E =$ Employment density (Total employment/Total households).

8. $\alpha =$ Ownership ratio (Traffic Survey estimates of total autos owned by households in 1963/Total Polk registrations in 1963).

Several automobile ownership statistics are used in the studies reported in this paper. None are ideal. R. L. Polk Co. statistics on total registrations in 1950 and 1960 are used in defining auto ownership per household. Polk registrations include both autos owned by households and those owned by businesses and other users, but the theory underlying the empirical estimation only refers to automobile ownership by households. Since business ownership per dwelling unit is almost certainly correlated with some of the explanatory variables, this might bias the parameter estimates. The ownership ratio is introduced as an explanatory variable in an attempt to reduce this source of bias.

Similarly, difficult problems are associated with the concept of residential density. In this framework, residential density (net dwelling unit density and percent single family) is interpreted as a measure of household or family consumption of residential space. Dwelling unit density is probably preferable on a priori grounds, since postwar residential development has been marked by both increased occupancy of single family units and larger lot size. However, since estimates of land devoted to residential use are fairly inaccurate for 1950 both measures are presented.[5]

While these problems of measurement are serious, variables not included in the analysis are potentially an even greater source of bias. Nearly everyone agrees the availability and quality of public transit affects the ownership of private automobiles and many would contend the density of residential development is similarly affected by the amount and kinds of available transit services. Yet for many reasons, not the least being measurement problems, no transit service variable is included in the statistical analysis. It is hoped that this omission may be at least partially corrected by extensions of the research reported here.

[5] While there was a complete land use survey in 1963, the amounts of land in residential use in 1952 were estimated from aerial photographs. This method provides fairly accurate measurements of the amount and density of new development, but it largely ignores changes within built-up areas. Thus, the estimates of net dwelling unit density in 1960 are probably reasonably accurate, but those for 1950 are much less reliable.

Additionally, changes in metropolitan employment distributions have affected both the competitiveness of public transit and the costs of lower density residential services. Much of the observed increases in automobile ownership and declines in lower residential densities in Boston, as elsewhere, must be due to employment dispersal.[6] Public transit is obviously much more competitive in dense central areas. Moreover, high residential land costs in central areas or the expense of commuting to cheaper peripheral land inhibits many centrally employed workers' from living at lower densities. If their jobs shift to suburban locations near cheap land or if they purchase a private automobile for commuting to work, the costs of residing at lower density will decline. Employment density, included in the auto ownership equation, may account for some effects of changes in the spatial distribution of employment. However, it is included primarily as a measure of the non-household demand for auto ownership.

Empirical Findings: The First Hypothesis

Both additive (arithmetic) and multiplicative (log-log) equations were estimated, but only the latter are reported. This choice was made primarily because the regression coefficients of multiplicative models have the convenient property of being constant elasticities. In addition, the multiplicative models were slightly better in terms of the proportion of explained variance. In any case, the signs, general magnitudes, and statistical significance of the regression coefficients were

TABLE 1

LEAST SQUARES ESTIMATES FOR RESIDENTIAL DENSITY: 54 BOSTON
COMMUNITIES IN 1950 AND 1960

Variables	Net Dwelling Unit Density		Percent Single Family	
	1960	1950	1960	1950
A	−1.07†	−2.92*	1.80*	1.90*
Y	−1.00†	1.85*	−0.47‡	−0.74†
F	−5.36*	−2.04‡	3.22*	1.84†
L	3.08*	3.26*	−2.49*	−3.21*
α	−0.90‡	−1.43†	0.94†	0.85‡
Constant......	18.6*	12.3*	−7.60*	−4.88*
R^2	0.79	0.75	0.84	0.74

* Coefficient differs from zero at the 0.01 level of significance.
† Coefficient differs from zero at the 0.05 level of significance.
‡ Coefficient greater than one standard error.

highly consistent for both formulations.

Table 1 summarizes least squares estimates based on the first hy-

[6] Rough calculations indicate that employment within 5 miles of downtown Boston declined from 65 to 57 percent of all Metropolitan area employment between 1950 and 1960. More significantly, the number employed in that area fell by 29,000 or by 6 percent.

pothesis: that residential density depends on auto ownership. Each measure of residential density is postulated to depend on income, family size, labor force participation per household, the ownership ratio, and auto ownership per household. Family size and labor force participation are included as two indicators of household preferences for lower density residential services. Larger families would generally be expected to prefer larger houses and lower densities. By comparison, multiple wage earner households and especially those where the housewife is the secondary wage earner might well find this extra space a burden or at least less desirable. The ownership ratio is included in the structural equations as an index of nonhousehold ownership of private automobiles.[7]

Both the signs and magnitudes of the regression coefficients (elasticities) are generally consistent as among the four equations; i.e., two dependent variables in 1950 and 1960. Of the four, the elasticities in the 1950 dwelling unit density equation appear most dissimilar, a hardly surprising result given the large measurement errors of that dependent variable in 1950.

Automobile ownership has the correct sign and an elasticity greater than one in all four equations. In the percent single family equations the estimated elasticities are nearly identical and much closer to two. For the dwelling unit density equations, the automobile ownership elasticities had identical signs, but that for 1960 is nearly three times as large as that obtained in the 1950 equation.

The family size and labor force participation variables appear to be the most important determinants of residential density, at least as measured by the magnitude, statistical significance, and consistency of the estimated elasticities. Both of these variables exhibit parameter shifts favoring lower residential density between 1950 and 1960 for both dependent variables. For example, in the percent single family models, for which the measure of residential density might be expected to be most consistent as between the two years, the elasticity of family size increased from 1.8 to 3.2, while that for labor force participation decreased in absolute value from −3.2 to −2.5. The large elasticities of the family size and labor force participation variables combined with the suggestion of parameter shifts between 1950 and 1960 suggests household preferences have been a powerful force in determining the form of postwar metropolitan development.

[7] Automobile ownership and the automobile ownership ratio are best thought of as a composite variable where the purpose for including the latter is to correct the automobile ownership coefficient for nonhousehold ownership. This method of correcting for bias is admittedly crude. However, analyses of a subsample of 34 cities of over 25,000 population, for which both 1960 Polk and 1960 census car ownership statistics were available, indicate this correction improves the parameter estimates of both residential density and automobile ownership equations.

Income appears to be both a less important and less consistent determinant of residential density. It has the wrong sign in three out of four of the least squares equations. Significantly, income has the correct sign only in the 1960 dwelling unit density equation. These findings reinforce the view that dwelling unit density (when accurately measured) is a better index of residential density than percent single family.

Parameter estimates obtained for the auto ownership ratio are also highly consistent as among the four equations. Since the ratio is measured for 1963, it should be less accurate for 1950 and should therefore reduce the bias by less in that year. However, since its sign is correct in both 1950 equations, it is presumed that it operates in the right direction.

Of the residential density equations, that for dwelling unit density in 1960 appears to provide the best overall estimates. In that equation the elasticities of both auto ownership and income are statistically different from zero at the 0.05 level, have the correct signs, and are approximately minus one.

Empirical Findings: The Second Hypothesis

Table 2 summarizes least squares estimates based on the second hypothesis: that lower residential densities, higher incomes, and larger families cause higher levels of automobile ownership. The empirical estimates obtained for the automobile ownership equations are, if anything, better than those presented above for residential density. The findings suggest that while residential density and family size affect automobile ownership, they are much less important than income. The estimated income elasticities are highly significant in all four equations

TABLE 2

LEAST SQUARES ESTIMATES FOR AUTOMOBILE OWNERSHIP: 54 BOSTON
COMMUNITIES IN 1950 AND 1960

Variables	Automobile Ownership			
	1960	1950	1960	1950
S............	—	—	0.19*	0.26*
D............	−0.08*	−0.20*	—	—
Y............	0.65*	0.64*	0.61*	0.54*
F............	0.44†	0.37	0.09	0.54‡
E............	−0.03‡	0.01	−0.01	0.02
α............	−0.43*	−0.17	−0.42*	−0.07
Constant.......	2.48*	2.30*	2.14	0.94
R^2..........	0.83	0.74	0.88	0.67

* Coefficient differs from zero at the 0.01 level of significance.
† Coefficient differs from zero at the 0.05 level of significance.
‡ Coefficient greater than one standard error.

and exceed 0.6 in three out of four, being only slightly smaller (0.54) in the 1950 single family equation.

The residential density elasticities also have the correct sign and are highly significant in all four equations. They range in absolute value from 0.08 to 0.26. The smallest absolute value is obtained for the 1960 dwelling unit density equation, where a 1 percent decline in density is associated with less than a one-tenth of 1 percent increase in car ownership. While family size has the correct sign in all four equations, it is statistically significant (at the 5 percent level) only in the 1960 dwelling unit density equation. The family size elasticities for the two 1960 equations are very different, being 0.44 and statistically significant at the 0.05 percent level in the dwelling unit density equation, but only 0.09 and less than one standard error in the percent single family equation. The absolute values of the residential density elasticities exhibit just the opposite ordering, being more than twice as large in the percent single family equation.

While the ownership ratio coefficients are nearly identical and highly significant in both 1960 equations, they are much smaller and less than one standard error in both 1950 equations. Yet they have the correct sign and are therefore retained. The employment ratio has a negative sign in both 1960 equations and a positive one in both 1950 equations. One plausible interpretation of this result would be that the ownership ratio adequately accounts for business auto ownership in 1960 and that therefore the employment density variable is either proxying better transit service or a similar characteristic of higher density workplaces. However, in 1950 when the ownership ratio variable is less adequate, employment density proxies business ownership.

These findings for the second hypothesis, suggest that family size and residential density only weakly affect automobile ownership, and that income is a far more important determinant. By comparison, the estimates in Table 1 indicate that family size and labor force participation are more important determinants of residential density, that income is only a weak and inconsistent determinant of residential density, and that it operates mostly through its effect on auto ownership.

Empirical Findings: The Third Hypothesis

The third hypothesis states that automobile ownership and residential density are jointly and simultaneously determined.[8] This view argues that a change in the level of either auto ownership or residential

[8] This by no means exhausts the causal hypotheses about the relationship between auto ownership and residential density. Several recursive structures could be postulated. For an econometric model using one such recursive structure see John F. Kain, "A Contribution to the Urban Transportation Debate: An Econometric Model of Urban Residential and Travel Behavior," *Rev. of Econ. and Statis.*, Feb., 1964.

density will affect the other and that changes in the original variable may be caused by either common underlying variables or by variables related to only one of them. For example, an increase in income might affect both residential density and automobile ownership directly and might affect them indirectly through its impact on the remaining variable as well. By contrast, a decrease in the price of automobiles or of gasoline would be expected to have a direct measurable effect only on automobile ownership, but it might affect residential density indirectly through automobile ownership.

Where variables are jointly and simultaneously determined in this way, use of ordinary least squares yields biased and inconsistent estimates of the parameters. The difficulty arises because the disturbance or error term and the endogenous explanatory variables are correlated. Two-stage least squares (TSLS) is one of the simultaneous equation estimation techniques that have been devised to deal with this problem.[9]

Table 3 presents TSLS estimates for residential density and auto

TABLE 3

SIMULTANEOUS EQUATION ESTIMATES FOR RESIDENTIAL DENSITY AND AUTOMOBILE
OWNERSHIP: 54 BOSTON AREA COMMUNITIES IN 1950 AND 1960

Variables	Net Dwelling Unit Density		Percent Single Family		Automotive Ownership			
	1960	1950	1960	1950	1960	1950	1960	1950
S........							0.19*	0.10
D.......					−0.14*	−0.11‡		
A.......	−2.26*	−0.77†	1.22*	1.42†				
Y.......	§	§	§	§	0.52*	0.62*	0.63*	0.57*
F........	−4.20*	−5.41†	3.80*	3.00†	§	§	§	§
L........	2.53†	3.75*	−2.75	−3.35*				
E........					−0.03‡	−0.05	−0.02	−0.06
α........	−1.37†	−1.61‡	0.72‡	1.14†	−0.45*	−0.09‡	−0.43*	−0.01
Constant.....	19.70*	14.11*	−7.03*	−5.74†	3.72*	2.49†	2.31*	1.92†
R^2.......	0.77	0.30	0.75	0.41	0.81	0.84	0.81	0.34

  * Coefficient differs from zero at the 0.01 level of significance.
  † Coefficient differs from zero at the 0.05 level of significance.
  ‡ Coefficient greater than one standard error.
  § Variable omitted because of multicollinearity problems.

ownership in 1950 and 1960. While problems of multicollinearity are present in some of the earlier estimates, they do not appear to cause serious difficulties. This is not true for the TSLS, however. Serious problems of multicollinearity exist for both the automobile ownership and residential density equations.

In the four residential density equations it is impossible to separate the effects of income and auto ownership. When income is added to the two 1960 residential density equations containing auto ownership, it

  [9] A discussion of two-stage least squares may be found in: J. Johnston, *Econometric Methods* (McGraw-Hill Book Co., 1963), pp. 258-63.

causes the sign of the auto ownership variable to change and to become nonsignificant. Auto ownership has the correct sign in the 1950 residential density equations but that of income is incorrect.

In the four automobile ownership equations the problem is serious collinearity between family size and residential density. It is impossible, with these samples, to separate the effects of family size and residential density on automobile ownership.

Given these problems of multicollinearity, it is necessary to omit either income or auto ownership from the residential density equations and either family size or residential density from the auto ownership equations in order to obtain meaningful parameter estimates. While omission of income and family size solves the collinearity in a technical sense, it is an unsatisfactory solution. The remaining variables, auto ownership and residential density, can only be said to reflect the combined effects of both variables.

No major changes occurred in the elasticities of the remaining explanatory variables used in the TSLS residential density and auto ownership equations. In the TSLS residential density equations, the preference variables appear to be even more important than in the parallel least squares equations and this tendency is especially marked in the case of family size in the 1950 equations. Also the coefficient of the ownership ratio is somewhat larger in the TSLS residential density equations. Changes in coefficients are even less marked for the TSLS auto ownership equations. The most notable change appears to be that the sign of the employment density variable is negative in both 1950 and 1960.

Backcasts and Forecasts

What are perhaps the most interesting results of this study were obtained when the several models were used to make forecasts and backcasts. Forecasts for the simultaneous equation model were obtained from the estimated reduced forms. The results of these experiments are summarized in Table 4. Comparisons are presented in terms of the coefficient of determination (R^2) and the mean error (ME), but the results for other measures are similar.

Backcasts for 1950 using the 1960 least squares (LS) equations are very good. Those using the 1960 estimated reduced forms (RF) are less good, with the 1960 net dwelling unit density reduced form doing an especially poor job of backcasting 1950 net dwelling unit density. This result is hardly surprising given the previously discussed errors in that dependent variable.

The forecasts for 1960 are somewhat more surprising. In five out of seven cases, including all three reduced forms, better fits are obtained using 1960 data (forecast) than 1950 data (actual). Moreover, while

TABLE 4

FORECASTS AND BACKCASTS OF RESIDENTIAL DENSITY AND AUTOMOBILE OWNERSHIP
EQUATIONS: 54 BOSTON AREA EQUATIONS IN 1950 AND 1960

Equations	Backcast: 1960 Equations			Forecast: 1950 Equations		
	Coefficient of Determination		ME	Coefficient of Determination		ME
	1960	1950	1950	1950	1960	1960
Net dwelling unit density						
1. LS..........	0.79	0.50	−0.74	0.75	0.56	−0.10
2. RF..........	0.80	0.15	−1.13	0.31	0.50	−0.05
Percent single family						
3. LS..........	0.84	0.72	−0.03	0.74	0.77	0.06
4. RF..........	0.76	0.34	0.34	0.42	0.59	0.07
Auto ownership						
5. LS with D....	0.83	0.57	0.06	0.74	0.64	−0.06
6. LS with S....	0.88	0.62	0.11	0.67	0.79	−0.02
7. RF..........	0.82	0.30	0.16	0.37	0.75	−0.03

the 1950 reduced form estimates all fit the 1950 data rather poorly,
their forecast accuracy is very high. These rather surprising results
and the generally good forecasting performance of the 1950 models
suggest that the serious measurement errors for 1950, discussed pre-
viously, are largely random. While these random errors lead to poorer
statistical fits, they do not seriously bias the parameter estimates.

Some Additional Evidence

Further evidence about the relationship between residential density
and automobile ownership is provided by two unpublished studies by
the author. Nearly identical models were estimated for a cross-section
sample of 286 London Traffic Districts using 1961 data. The London
study includes least squares and TSLS estimates for both automobile
ownership and residential density (measured by net dwelling unit den-
sity). Least squares estimates for auto ownership were also obtained
from cross-section analyses of 1960 census data for the central cities
and suburban fringes of the sixty-five largest U.S. urbanized areas.
These intercity analyses of automobile ownership employed a some-
what different set of explanatory variables than are used in this paper.

Income and density elasticities obtained in these two studies are re-
markably consistent with those reported above. This consistency is
especially marked for the income elasticities obtained for the least
squares auto ownership equations. For the Boston study the income
elasticity was about 0.6, for the suburban fringe sample about 0.4, for
the central city sample about 0.5, and for the London sample about
0.7. Most persons asked to order these four samples according to

timing of development, level of income, and auto ownership, would place Boston between the central city and the London samples. Thus, the income elasticities are not only amazingly similar, but they have a tantalizing order.

Similar, if somewhat less regular, results were obtained for residential density in the automobile ownership least squares equations. Residential density elasticities for the four Boston study automobile ownership least squares equations were themselves less consistent, ranging in absolute value between 0.09 and 0.26. (Absolute values are used in discussing the residential density elasticities since some are defined in terms of density and some in terms of the reciprocal of density. All have the correct signs.) However, three of the four values fell between 0.19 and 0.26. The residential density elasticity for the central city sample was −0.18, that for the suburban fringe sample about 0.25, and that for the London study about −0.36. The magnitudes of the residential density elasticities have a somewhat different order than those for income, but the results are remarkably similar given the differences in variable definition and model specification.

Findings for the London TSLS estimates were also strikingly similar to those obtained for Boston. Identical multicollinearity problems arose in the TSLS estimation. In the residential density equation income and automobile ownership were so highly collinear that income had to be omitted. Similarly, family size and residential density were so collinear in the auto ownership equation that the family size had to be omitted.

Conclusions

While this paper does not provide clear-cut or conclusive answers about the interrelationship between residential density and automobile ownership, it does provide much new and consistent information. Regardless which causal hypothesis is accepted, it appears that income has been the most important factor underlying both higher postwar levels of automobile ownership and declines in residential density. This conclusion derives both from consideration of the estimated regression coefficients and of the postwar changes in income and other explanatory variables.

Real median family income in the Boston SMSA increased by an estimated 56 percent between 1950 and 1960. The income elasticities obtained for the least squares auto ownership equations (Table 2) indicate that an increase of this magnitude in real median family income would increase auto ownership per household by between 28 and 34 percent. By comparison, the elasticities obtained for the percent single family variable indicate that the decade's 39 percent increase in the proportion of Boston households living in single family units would account for only an 8 to 10 percent increase in auto ownership per house-

hold. Moreover, while residential density is exogenous in the least squares formulation, it is at least partially dependent upon income. Thus, part of this 8 to 10 percent increase in auto ownership might also be attributable to the growth in income during the decade.

The multicollinearity between auto ownership and income in the TSLS estimation makes it difficult to interpret the simultaneous equation results. Even so the parameter estimates obtained for the TSLS equations suggest the single equation models provide conservative estimates of the effect of increases in income on auto ownership.

It also appears that rapid postwar increases in family income have strongly affected residential density, either directly as an item of consumption or indirectly through their effect on automobile ownership. The 1960 net dwelling unit density equation seems the most satisfactory of the four least squares residential density equations, by far. In it the income elasticity is approximately minus one indicating that the 56 percent increase in real median family income between 1950 and 1960 would cause about a 56 percent decrease in net dwelling unit density. The automobile ownership elasticity in the same equation is also approximately minus one indicating that the estimated 29 percent increase in auto ownership per household during the decade should have caused about a 29 percent decline in net dwelling unit density. While auto ownership is an exogenous variable in this formulation, it would appear to be strongly dependent on income. Thus, part of the 29 percent decline in net dwelling unit density due to increased auto ownership must be attributed to rising income.

Multicollinearity problems in the TSLS estimation also create problems in interpreting the importance of income as a determinant of postwar declines in residential density. However, the TSLS results again suggest that the least squares estimates of the effect of income on net dwelling unit density may be conservative.

In terms of contribution to explained variance and the magnitude and statistical significance of their regression coefficients, family size and labor force participation appear to be even more important determinants of residential density. However, they are much less powerful as explanations of postwar changes in residential density during this period, since changes in them have been much smaller than increases in income.

While the models presented here appear to provide surprisingly consistent results for a fairly wide range of samples, their partial character is painfully obvious. The omission of transit service variables and inadequate specification of employment location are only the most obvious examples. Specification and measurement of these concepts will not be easy tasks, but they would appear to be essential if the processes of metropolitan development are to be fully understood.

DISCUSSION

JULIUS MARGOLIS: In reading these three papers one is again impressed by the great difficulties of analyzing the structure of activities in the metropolitan area. It is clear why so few economists have persisted in these studies. The economist, as a theorist, is prone to simplify; i.e., to concentrate on the minimum number of factors necessary to explain a large system of interactions. The economist, as an empirical tester of hypotheses, is rarely satisfied with a partial answer, a simple relationship. He reaches out for links with unexplained parts of the economy. Unfortunately, the urban structure is a frightfully complex set of mutual interdependencies which on the one hand have made the typical economists' models seem very oversimplified and on the other hand seem to make partial empirical analysis too restrictive. The three papers are excellent examples of the two approaches. Written by the most eminent practitioners, they demonstrate both the strengths and difficulties of the economist in dealing with urban form.

I shall restrict my comments to the Mills paper, though my observations are shaped by the studies of Moses-Williamson and Kain. Unlike the authors, I am a consumer of models of urban structure rather than a producer, and, therefore, my comments may have the irresponsibility of the frustrated user who feels confident that a better product could be built, but all he may be demonstrating is his ignorance of the technology.

Complex interdependencies are no mystery to economists. After all, this is the essence of general equilibrium analysis. There are at least three difficulties in analyzing urban form via economic models. The Mills model, despite its many virtues, does not face up to these difficulties. Externalities rather than market interdependencies permeate the urban area. Traffic generation and congestion are the most dramatic forms of externalities, but externalities appear in many other forms as well. How are they to be handled?

Externalities are often priced out through the land market. Nonmarket links among firms and among firms and households are often managed by locational decisions and, thereby, through the market for sites. But the urban land market is a most imperfect market. What does the model assume about the operations of the land market?

One of the major impediments in site adjustment is the long life of buildings, bridges, streets, parks, etc. Few urban econometricians lament the absence of quarterly or even annual data to estimate the parameters of their models of urban form. Though it is standard to stress the process of adjustment in analyzing market behavior, it is difficult to visualize a reasonable approximation model for an urban area which does not consider the long-continuing influence of old technologies and past allocations of capital, land, and population to previous forms. How does time and life of structures enter into the model?

Mills states that his model is simplified and aggregative, and though he is correct, it is less simple and aggregated than a typical equilibrium market

model, but possibly it is overly simplified and overly aggregated for a model of urban form. Simple models are useful. Certainly some of his results are of great interest. The finding that if we analyzed the central business district cost function as a long-run planning curve, we would not run out of land; there is no necessary congestion; and that we would have to look elsewhere for sources of congestion are extremely useful building blocks. Simple models are valuable and, unfortunately, all my reservations involve complications of the model which might make it unmanageable.

Mills deals with a monocentric city. Business subject to external economies of scale is located downtown, and activities with constant returns to scale are located in the suburbs. But we know this is not so. Not only are there many nodes of economic activity, but centers of accessibility differ for different activities. Urban centers, defined by minimizing air mile distance, tell us too little. Transportation costs are what we are after, and transportation costs are very different for workers, goods, shoppers, salesmen, and executives. Certainly the changing ratio of cost of moving persons to cost of moving goods has led to serious changes of urban form as suggested by Moses-Williamson. But Mills does not offer us enough insight into the production process of transportation or the functional relationship between transportation costs and location to help us move beyond the monocentric model. The candidates for those who are included in the central business district are dictated by external economies of scale where presumably transportation and communication economies are most significant; but these are not explicit, and possibly an explicit treatment would help us move to a less rigid monocentric model. The economies of transportation, communication, and of specialization in production are implicit in a central business district increasing returns production function. The central business district is then handled as a unit, and since we know that monopoly follows from increasing returns, Mills assumes a market situation where the central business district sells, as a monopolicy, on the world market. This seems to be a strange and unnecessary complication to introduce. The increasing returns are in the supply of business services, transport, and information links rather than production processes of the exporting firms. Each firm may be operating as a competitor in the world, or even within their metropolitan markets, even though there are agglomeration economies to the entire package of firms.

The aggregation assumptions of the central business district are matched by similar necessarily simplifying statements in regard to other parts of the city. For instance, the issues of concern to Kain which deal with the location of residences within the urban area are absent. There are no utility functions of households; family size, income, job opportunities, density preferences are all absent. There are no intractable building sites, no fixed street patterns, and no cross-hauling of goods and passengers.

It is patently unfair to charge Mills with neglect of these factors. He knows their relevance far better than I, and I am sure that he toyed with their introduction as well as considering still other factors. That he did not deal with them is a function of a predilection to try to exploit the models which have

been developed in economics, a necessary surrender to the difficulties of ana-
lytical solutions when this model becomes too cumbersome, and a willingness
to aspire to explain a nineteenth-century city rather than a metropolitan area.
Within these constraints he has done a masterful job of extending the work
that has been initiated in the last decade. Of concern to me is whether to en-
courage this type of model. I suspect that actual urban facts are too far from
the idealized urban facts which he seeks to explain and, therefore, a firm ad-
herence to a microeconomic model quickly becomes cumbersome. If both of
these concerns are appropriate, then other analytical models may be neces-
sary. Computer simulation may be preferable; possibly more partial analysis
may be desirable. A general analytical spatial model is certainly useful, but we
may not yet have the appropriate technology for it.

BENJAMIN CHINITZ: The authors stress the role of changes in the structure
of transport costs on the location of manufacturing in the metropolitan area.
They feel that inadequate attention has been given this force in explaining
the changing pattern of location. This is a claim which is difficult to evaluate
in a general and conclusive sort of way, but I must immodestly call attention
to my own work on the New York area, *Freight and the Metropolis*, in which
I devoted a whole chapter to precisely this point. Happily, our conclusions are
consistent. I found in looking at the changing distribution of employment
between city and suburb that those industries in which the movement of
freight is a significant component of cost, show a greater tendency to move
away from the center.

I would insist, however—and I am sure that the authors would not dispute
this—that the transport cost hypothesis cannot stand alone as an explanation
of shifts in the location of manufacturing within metropolitan areas. More-
over, the transport cost hypothesis itself has aspects and dimensions which
were not considered by the authors. Let me consider these first.

I have in mind particularly the changing technology of freight handling as
opposed to freight movement. The savings resulting from proximity to interci-
ty terminal facilities not only had to overcome higher labor and land costs but
also the inefficiencies induced by the need to handle freight in cramped quar-
ters. Thus, even under the old technology, labor productivity in freight han-
dling must have been lower in the close-in areas than in the more distant
areas. But the margin was surely widened dramatically when technology
opened up opportunities to substitute electric power for human muscle, in the
form of conveyor belts, fork lift trucks, etc. The full exploitation of these op-
portunities called for the use of more ground space and would in itself, almost
independent of the truck, have favored the relocation' of manufacturing in
more remote areas.

It may not even be necessary to invoke technological change at all to ex-
plain the drift to the suburbs. The authors have cast the location decision in
terms of the relative cost of transportation and other factors, particularly
labor. One can also view the choice in the most elementary terms of intro-
ductory production theory.

Location is a variant along the output contour. More capital and less labor is associated with greater use of space and hence a preference for suburban or further out locations. More capital and less labor is associated with less space and close inward locations (ground space). The choice then depends on the relative cost of capital and labor which we can assume to be equal in all parts of the area. (The authors would not like this because it is critical to their argument that labor is more expensive at the center.)

The slope of the price line is in turn a function of supplies of factors. If we study the city against the backdrop of an increasingly affluent economy in which capital accumulation is a significant component of economic growth, we can infer rather directly that the core is threatened by the accumulation of capital which increases the cost of labor relative to the cost of capital.

My main point, however, is that it would be a mistake to minimize the role of external economies in explaining the concentration of manufacturing in the core, especially in cities like New York, Philadelphia, Boston, and probably also Chicago. One cannot explain the relative concentration of certain industries in the core with reference to the need to be close to terminals of intercity transport. On the contrary, at any moment of time, it is precisely those industries which generate little freight per dollar of value added which have the greatest affinity for the core. For industries like apparel, jewelry, novelties, etc., the location calculus clearly runs in terms of higher labor costs versus economies of agglomeration.

Curiously enough, however, the authors are right for the wrong reasons in calling attention to the role of transport costs even in industries of this kind. What the new technology makes possible is the searching out of labor costs for those aspects of the process which are standardized and amenable to mass-production techniques, while retaining the core location for the unstandardized functions of design and marketing. In other words, the truck has made it possible to disintegrate the operation locationally to take advantage of the best of both worlds. On the other hand, the search for lower labor costs often takes the manufacturer way beyond the suburbs to much more distant locations. With jet service, even Tokyo and New York City can be straddled in this manner. Clearly, though, this is a different kind of adaptation to changing transport costs than the authors had in mind.

RICHARD F. MUTH: I wish to congratulate the authors of the three truly fine papers presented at this session. Each paper is, in my opinion, an important contribution to the field of urban economics. Mills has accomplished

what few of us working in this field have been willing to tackle: the development of an analytically manageable, general equilibrium model of a city. The paper by Moses and Williamson is a most welcome exploration of an important but neglected problem: the determinants of the intracity location of manufacturing firms. Kain provides valuable empirical evidence of the determinants of the intensity of residential land use. Since my own interest is principally in urban residential land use, my remarks will be directed mostly to the papers by Mills and Kain, but I do not wish to imply by so doing that I find the third paper any less stimulating.

The implications of Mills's model which are most interesting to me are those relating to the decline in population densities with distance from the city center. As Mills notes, no special case of his model yields the negative exponential population density function, either in gross or in net terms. Despite the difficulties in evaluating Clark's results which Mills notes—plus the additional one that Clark's observations were for annular areas surrounding the city center and few in number for any given city—the scatter diagrams Clark presented were indeed striking. More recently I have made estimates of the relation of the log of gross population density to distance for each of 46 U.S. cities in 1950.[1] Too many statistically significant departures from log-linearity were observed to attribute to sampling variation. But, among the coefficients of the quadratic terms in distance, neither positive nor negative signs tended to predominate. I have also examined the logs of gross and net population densities in relation to distance and other variables for census tracts on Chicago's south side.[2] Both for 1950 and for 1960 gross and net population densities declined at about the same relative rate. For neither year did the addition of a quadratic term in distance in either the gross or net density regression improve the fit very much, though the curvature was positive in the log net density regression for each of the two years, as Mills's equation (24) would imply. Thus, I believe there is enough evidence that the negative exponential density function approximates the actual pattern that it should not be discarded until some other is shown to be better.

Supposing that one would want to obtain negative exponential population densities, how should Mills's model be modified? The best way, I believe, would be to replace the Cobb-Douglas housing production function Mills uses, his equation (3), though I realize that dropping it would greatly complicate the solution to Mills's model. Since in the model housing consumption per worker is constant, the output of housing per unit of land is the only component of net population density which varies with distance. Assuming constant returns to scale in producing housing, as Mills does, it can be shown that

$$(1) \qquad \frac{d}{du} \ln \left(\frac{X_3}{L_3} \right) = \left\{ \frac{(1 - \rho_L)}{\rho_L} \sigma \right\} \left(\frac{p_3'}{p_3} \right),$$

---

[1] "The Spatial Structure of the Housing Market," *Papers and Proc. of the Reg. Sci. Asso.*, 7 (1961), pp. 207-20. For each city a random sample of 25 census tracts outside the CBD was used, the latter being that identified by the 1954 Census of Business.

[2] "The Variation of Population Density and Its Components in South Chicago," *Papers and Proc. of the Reg. Sci. Asso.*, 15 (1965), pp. 173-83.

where, in addition to the notation Mills uses, ρ_L is total payments to land relative to the value of output and σ is the elasticity of substitution of land for other factors in producing housing. From Mills's equations (9), (13), and (14),

$$(2) \qquad \left(\frac{p_3'}{p_3}\right) = - (a/x_3 \overline{A}_3^{1/\alpha 3}) p_3^{(1/\alpha 3)-1}.$$

Hence, for $\alpha_3 < 1$, housing prices decline at a numerically declining relative rate with distance. Since for the Cobb-Douglas function $\sigma = 1$ and ρ_L is constant, the coefficient of (p_3'/p_3) in (1) is a constant. Hence, (X_3/L_3) and net population densities decline at numerically smaller relative rates at greater distances. However, with a less than unit elasticity of substitution in producing housing, ρ_L would decline with distance as land rentals do. Hence, the coefficient of (p_3'/p_3) in (1) would increase with distance and offset the latter's numerical decline. Some evidence for a less than unit elasticity of substitution is provided by the fact that in the postwar period the ratio of site to house values has risen as land rents have risen relative to construction costs.[3]

The specific housing production function he uses is also responsible for Mills's conclusion that, for $\alpha_3 < 1$, land rents in the "suburbs" must decline at a numerically declining relative rate. It can also be shown that

$$(3) \qquad \left(\frac{R'}{R}\right) = \frac{1}{\rho_L}\left(\frac{p_3'}{p_3}\right)$$

if the profits of housing producers are to be everywhere the same. With a less than unit elasticity of substitution production function, the decline in ρ_L with distance would tend to offset the decline in$(-p_3'/p_3)$, so that a roughly constant relative decline in land rentals could be achieved. In Mills's model it is not merely the possibility of economizing on land, but of economizing sufficiently, to prevent residential land rents from rising exponentially as one approaches the city center.

Turning to Kain's paper, I would like to elaborate on his conclusion that demographic characteristics are much more important than automobile ownership in determining residential density. Such a conclusion is not surprising for parts of a given urban area, given the tendency for families of different types to distribute themselves unevenly within an area. As Kain notes, however, over time changes in automobile ownership may be much more important. Recently I estimated that the elasticity of the land area occupied by an urbanized area with respect to car registrations per capita is about 1.35, using data for 36 U.S. urbanized areas in 1950.[4] The elasticity of average population density is, of course, numerically the same but opposite in sign. This value is quite similar to Kain's estimate

[3] See my "The Derived Demand Curve for a Productive Factor and the Industry Supply Curve," *Oxford Econ. Papers,* New Series 16, July, 1964, p. 229.
[4] "The Distribution of Population in Urban Areas" (paper presented at the National Bureau of Economic Research Conference on Investment Behavior, June, 1965).

of net dwelling unit density with respect to auto ownership in 1960 in his Table 1. In 1950, the middle two-thirds of values of car registrations per capita ranged from 0.20 to 0.30, about 38 percent of the mean value of 0.26. According to the above noted estimate, such a variation would have produced differences of roughly 50 percent in average population density. More important still, from 1950 to 1960 the average density of the areas I studied declined by about 52 percent, while car registrations increased from 0.26 to 0.35, on the average for all areas. The latter by itself would account for a reduction in average population densities of about 47 percent —almost all the observed decrease over the decade. Thus, particularly over time, the effects of transport cost reductions as reflected in increased car registrations can indeed be of substantial practical importance.

ANTITRUST AND MONOPOLY

THE GOALS OF ANTITRUST POLICY

By ROBERT H. BORK

Yale University

· The life of the antitrust law—meaning by that its areas of policy growth—is, in contrast to Holmes's dictum about the common law, neither logic nor experience but bad economics and worse jurisprudence. The economics consists of a woefully unsophisticated theory of the means by which firms can gain monopolies, or at any rate injure the competitive process and so injure consumers, by attacking or foreclosing their rivals.[1] The jurisprudence, which is the topic of my paper, consists of the notion that under existing antitrust statutes the courts may properly implement a variety of mutually inconsistent goals, most notably the goals of consumer welfare and small business welfare. Together, these ideas are creating a broad trend of policy directed less to the interest of consumers in free markets than to the interest of inefficient producers in safe markets. The question is whether this trend is either necessary or proper. I think it is neither.

The subject has been debated in a variety of ways. It may be helpful for a lawyer to argue the neglected point that the larger goals of antitrust policy are necessarily affected, in fact determined, by the institutional arrangements we have chosen to implement the policy.

My thesis is that existing statutes can be legitimately interpreted only according to the canons of consumer welfare, defined as minimizing restrictions of output and permitting efficiency, however gained, to have its way. That much I think is required under present statutes by the nature of the judicial process. You will notice, however, that the implications of my argument sometimes overflow the bounds of existing institutions and would support the view that no revision of antitrust should, as a question of legislative wisdom, depart from the consumer welfare premise.

If I am correct, reform is needed, but it need not come from Congress. Antitrust policy is determined, far more than most people realize, by the Supreme Court. Reform is as likely to come through change in the intellectual world which ultimately reaches the Court as by any

---

[1] The controlling article is Director and Levi, *Law and the Future: Trade Regulation,* 51 *Nw. U.L. Rev.* 281 (1956). See also Bork and Bowman, *The Crisis In Antitrust,* 65 *Colum. L. Rev.* 363 (1965), and *Contrasts in Antitrust Theory: I* and *II, id.* at 401 and 417.

other means. Reform by the Supreme Court, moreover, is more likely to achieve clean theoretical lines than reform achieved through the political process. The existence of an unelected, somewhat elitist, and undemocratic judicial institution thus makes theory more important than it might otherwise be in our governing processes. I am not suggesting a judicial coup d'etat. Rather, I intend to argue that an exclusive adherence to a consumer welfare test is the only legitimate policy for the Supreme Court under present statutes precisely because of the Court's elitist, unrepresentative nature.

Economics may be a science best confined to analysis and shunning the normative. The lawyer who ignores norms, however, is likely to descend to the level of gossip. Lawyers are properly concerned, not merely with the ways legal institutions behave in fact, but with models of how they ought to behave and how their behavior can be brought closer to the model. The time has come for lawyers to lessen their traditional preoccupation with what courts do in fact and to construct normative models of judicial behavior. Courts as institutions are not fully subject to the discipline of either an economic or a political marketplace. We can attempt to remedy that by constructing an intellectual marketplace which disciplines judicial power through informed criticism. For that task we need a model from which deviations can be measured. This paper attempts to state part of the model appropriate for courts enforcing the antitrust laws.

Social values other than consumer welfare have always played a part in, at least, the rhetoric of antitrust. As early as 1897 the Supreme Court expressed concern over combinations that reduced prices through efficiency, saying, "trade or commerce . . . may . . . be badly and unfortunately restrained by driving out of business the small dealers and worthy men whose lives have been spent therein, and who might be unable to readjust themselves to their altered surroundings."[2] More recently Judge Learned Hand informed us that "one of [the] purposes of [the antitrust statutes] was to perpetuate and preserve, for its own sake and in spite of possible cost, an organization of industry in small units which can effectively compete with each other."[3] He believed that social and moral considerations could properly override consumers' economic welfare. These strains in the law have grown and prospered. They are virulent in decisions under the Robinson-Patman Act and the antimerger statute, amended section 7 of the Clayton Act. In the Von's Grocery case[4] a majority of the Supreme Court was willing to outlaw a merger which did not conceivably threaten consumers in order to help

[2] *United States* v. *Trans-Missouri Freight Ass'n,* 166 U.S. 290, 323 (1897).
[3] *United States* v. *Aluminum Co. of America,* 148 F.2d 416, 429 (2d Cir. 1945).
[4] *United States* v. *Von's Grocery Co.,* 384 U.S. 270 (1966).

preserve small groceries in the Los Angeles area against the superior efficiency of the chains.

Consumer welfare is the only legitimate goal of antitrust, not because antitrust is economics, but because it is law. At the risk of provoking hilarity in Professor Director, I invoke the cliché that the law has a life of its own. I appeal to three interrelated aspects of the legal process which most of us assume:

1. The functions of courts and legislatures are different, the latter making major political choices. Courts inevitably make policy, but where Congress has written a statute, the policy movements of the courts are ideally molecular rather than molar.

2. An important function of the courts, performed in a variety of ways, is to help keep the legislative process responsible by ensuring, so far as possible, that major policy decisions by the legislature are deliberately and openly made.

3. The judicial process itself must be responsible. That requires the decision of cases upon criteria which are judicially administrable, give fair warning to those required to obey the law, permit sufficient predictability so that desirable conduct is not needlessly inhibited, and permit rational explanation of the application of the criteria so that judicial performance may be evaluated and controlled.

The first two principles imply that it is the task of the courts under the antitrust laws to determine what policy choice the Congress has made and to apply it, but, if the Congress has not made its choice clear, to press that institution to live up to its responsibility. An examination of the language and legislative history of the antitrust laws makes it clear that consumer welfare was at a minimum a primary intended value. For my argument I do not need to contend that it was the sole intended value, since I will attempt to show the impropriety both of any other single value and of multiple values.

The first place to look for the policy choice of the legislature is in the words of the statutes. The language of the Sherman Act tells us little, but every major antitrust statute after the Sherman Act speaks in economic terms. Section 7 of the Clayton Act makes illegal corporate mergers "where . . . the effect . . . may be substantially to lessen competition, or to tend to create a monopoly. . . ." Very nearly identical language is used in section 3 of the Clayton Act with respect to requirements contracts, exclusive dealing contracts, and the like. Even the Robinson-Patman Act prohibits price discrimination only "where the effect of such discrimination may be substantially to lessen competition or tend to create a monopoly . . . or to injure, destroy, or prevent competition with any person. . . ."

The polar models employed by these statutes are thus "competition"

and "monopoly." These are models developed by economists rather than sociologists and appear, therefore, to frame a law to which eco-·nomic analysis, and not much else, is relevant. As models, competition and monopoly indicate different outcomes for resource allocation. The preference for competitive rather than monopolistic resource allocation is most clearly explained and firmly based upon a desire to maximize output as consumers value it. The language of the statutes, then, clearly implies a consumer welfare policy.

The legislative history of the Sherman Act shows an intention that the courts implement only a consumer welfare policy.[5] Later statutes are less clear, some congressmen displaying both a consumer-welfare premise and an intention that the protection of small business be accomplished through antitrust, often without realizing the conflict. What should a court do? Suppose that the unfocused, sloganistic intent to help small business could be implemented only by a statutory interpretation which would have serious deleterious effects upon national wealth. Would any court really be justified in translating such vaporings into hard results probably never foreseen even by the rhetorician, and certainly not by many of his colleagues? Given the courts' function of assisting the legislature to perceive and face the choices that must be made, the courts can and should use the language of the antitrust statutes and the other aspect of legislative history to hold to a consumer welfare policy. This will force Congress, if it wishes to change that policy or create exceptions to it, to face the real issues, the benefits and costs of alternative policies, and to make the legislative choice.

Some people suggest that the legislative intent was not really unfocused, that Congress really intended to sacrifice consumers to small business but found it politically expedient to phrase the statutes in the language of competition. Courts, it seems to be suggested, should rely not upon the straight-faced statutory command but upon the discreet congressional wink. But the purpose of a wink is to indicate the opposite of what one is saying in order to deceive a third party who hears only the words. The third party to a dialogue between the Congress and the federal courts is the electorate. There is surely a strong case that the courts should take Congress at its word, and require that body, if it really wants contrary results, to phrase its law in words that make the political decision and its costs apparent, not only to the courts, but to the electorate and the legislators themselves. Otherwise the legislative process becomes something of a fraud, and the courts are accessories.

[5] My argument on this point will appear in an article entitled "Legislative Intent and the Policy of the Sherman Act," Vol. IX of the *J. of Law and Econ.*

The third aspect of the legal process—judicial responsibility—is by far the most important for my argument. Courts simply cannot achieve the elements of responsibility already mentioned if they decide each case according to an essentially unstructured choice between consumers and inefficient producers. There exists no social science, no set of criteria, which could guide the choice in the particular case. Courts could achieve some of the elements of responsibility (e.g., administrable criteria, predictability) by opting either to protect only consumers or to protect only inefficient producers. The latter choice is not legitimately open to the courts, ·however. It would run contrary, not only to the competition language of the statutes and the legislative history, but would require destruction of such long-standing features of antitrust as the rule of per se illegality for cartels. In effect, the courts, to be consistent, would have to outlaw all creations or expressions of efficiency and to sanction all cartels. No one thinks that is a legitimate interpretation of the goals underlying the antitrust laws. We are left, then, with a choice between strict adherence to a consumer welfare rationale or a case-by-case compromise between consumer and producer interests.

In my view, the desideratum of judicial responsibility is so crucial that even if Congress had written statutes which explicitly ordered the federal courts to balance consumer and small business interests in each case, which it did not, I think the courts should flatly refuse to accept the delegation. Indeed, if the situation were put that baldly, I think the courts would refuse. A court is not the proper institution, either by equipment, responsiveness to the electorate, or specialization of function, to write *ab initio* detailed specifications of political compromise between conflicting and incommensurable values. I am at a loss why anyone thinks courts should dictate the terms of a compromise in antitrust any more than they should have written the tariff laws, the code of labor relations, or chosen the rate of progression for the income tax.

My belief that the Court, if it recognized the issue, would refuse to make political choices between small business and consumers is reinforced by the history of the Court's behavior. Defendants in the first Sherman Act price-fixing case to reach the Supreme Court urged that they be judged according to the "reasonableness" of the price they fixed.[6] A reasonable price, according to important common law precedent, was one fair to both the participants in the cartel and to the public.[7] The formula was thus one for judicial mediation between

[6] *United States* v. *Trans-Missouri Freight Ass'n,* 166 U.S. 290 (1897).
[7] E.g., *Nordenfelt* v. *Maxim Nordenfelt Guns and Ammunition Co.* [1894] App. Cas. 535, 565. The application of the common law formula emasculated an Australian antitrust

conflicting producer and consumer interests. The Court declined this invitation to case-by-case legislation and instead adopted a rule of per se illegality.[8] As Judge Taft phrased it the following year, a reasonable-price test forces judges to "set sail on a sea of doubt"[9] with no guide "except the vague and varying opinion of judges as to how much, on principles of political economy, men ought to be allowed to restrain competition."[10]

In the case of criminal statutes, at least, the problem takes on constitutional dimensions. For example, the Court in *United States* v. *Cohen Grocery Co.*[11] struck down as void for vagueness section 4 of the Lever Act which provided "that it is hereby made unlawful for any person willfully . . . to make any unjust or unreasonable rate or charge in handling or dealing in or with any necessaries; to conspire, combine, agree, or arrange with any other person . . . (*e*) to exact excessive prices for any necessaries. . . ." The Supreme Court said, in terms precisely applicable to the suggested balancing of consumer, producer, and other social interests: "To attempt to enforce the section would be the exact equivalent of an effort to carry out a statute which in terms merely penalized and punished all acts detrimental to the public interest when unjust and unreasonable in the estimation of the court and jury."[12]

Similarly, in *Cline* v. *Frink Dairy Co.*[13] the Supreme Court held the Colorado Antitrust Act unconstitutionally vague because it made the lawfulness of certain conspiracies and combinations turn upon a determination of "reasonable profit." And in *United States* v. *Trenton Potteries Co.* the Court once more rejected the reasonable-price test, stating: "In the absence of express legislation requiring it, we should hesitate to adopt a construction making the difference between legal and illegal conduct in the field of business relations depend upon so uncertain a test as whether prices are reasonable—a determination which can be satisfactorily made only after a complete survey of our economic organization and a choice between rival philosophies."[14]

Yet, without express legislation requiring it, this is precisely what the Court does do every time it undertakes to decide a particular case by weighing the social value of small business against the value of increased wealth to consumers.

---

statute similar to the Sherman Act in *Attorney General of Australia* v. *Adelaide S.S. Co. Ltd.* [1913] App. Cas. 781.
   [8] 166 U.S. 290, 342 (1897).
   [9] *United States* v. *Addyston Pipe & Steel Co.*, 85 Fed. 271, 283-284 (6th Cir. 1898).
   [10] *Id.* at 283.
   [11] 255 U.S. 81 (1921).
   [12] *Id.* at 89.
   [13] 274 U.S. 445 (1927).
   [14] 273 U.S. 392, 398 (1927).

United States v. *National Dairy Products Corp.*,[15] ¿
demonstrates both that the Supreme Court continues to]
with the vagueness of a reasonable-price test and that it]
proaches than a declaration of unconstitutionality to empl
logue with Congress, in this case the device of statutory
The Court majority saved section 3 of the Robinson-'
which makes it a criminal offense to sell goods at "unre;
prices for the purpose of destroying competition or elimir
petitor," only by interpreting it to prohibit sales belo\
with predatory intent. This reading transformed a reasonal
requiring a compromise between inconsistent values, into ¿
sumer-oriented statute using cost and intent as tests. Th
ing function of the concept of predatory intent in antit
been to assure the courts—whether rightly or wrongly
present point—that a particular practice is not an e
efficiency but reflects merely the desire to gain monopol
consumers. A three-member minority was not satisfied w
tion to the statute's vagueness and voted to hold it void.

As a matter of history, the Supreme Court has emplo\
for-vagueness doctrine only where criminal statutes wer
But that certainly cannot be taken to indicate that the sar
criteria are unobjectionable where civil sanctions are invc
particularly true of antitrust where civil sanctions—tri\
divestiture, and broad injunctions—are likely to be at le¿
as the available criminal penalties.

Apparently the present Supreme Court has not fully p
any attempt to write political compromises between co
producers involves the judicial improprieties that the (
wisely avoided in rejecting the reasonable-price test. If
doubt that the problem is real, it should be laid to rest b\
Brown Shoe opinion holding illegal the acquisition of a ¿
chain by a shoe manufacturer, partly because of the an
verse impact upon smaller shoe retailers: "Of course, so\
sults of large integrated or chain operations are beneficial {
Their expansion is not rendered unlawful by the mere fa
independent stores may be adversely affected. It is con
competitors, which the Act protects. But we cannot fail
Congress' desire to promote competition through the prot
ble, small, locally owned businesses. Congress appreciate
sional higher costs and prices might result from the m¿
fragmented industries and markets. It resolved these co

[15] 372 U.S. 29 (1963).

siderations in favor of decentralization."[16] No matter how many times you read it, that passage states: although mergers are not unlawful merely because small independent stores may be adversely affected, we must recognize that mergers are unlawful when small independent stores may be adversely affected.

Given the fragmented nature of the shoe industry, particularly in retailing, the ease of entry, and the obvious fact that the acquisition was motivated by a search for increased efficiency, the Court must either have been saying that any efficiency-creating merger is illegal because it threatens less efficient rivals or that, for reasons it could not articulate, this particular merger threatened social values other than consumer welfare. The former statement is, of course, illegitimate. Except where a special neutral motive, such as a quirk in the tax laws, is operative, all mergers are presumably motivated either by a desire to increase efficiency or to gain a position making output restriction profitable. Mergers of the latter type are clearly illegal. If mergers of the former type are also illegal, then virtually all mergers are outlawed, and that is a law Congress plainly did not write.

But if the Court was not announcing the per se illegality of mergers, then the incoherence of this passage from Brown Shoe demonstrates that the introduction of producer protection goals into antitrust subverts judicial virtue and responsibility. I believe no one could successfully maintain that the "criteria" laid down by the quoted passage are judicially administrable, give fair warning, avoid the needless inhibition of desirable mergers, or fulfill the Court's obligation, arising from its elitist, unrepresentative nature, to demonstrate the reasoned derivation of its decisions from known premises.

The requirements of space permit me merely to touch upon a variety of objections which are sometimes raised to my argument. One arises from the observation that courts often behave in the way I have characterized as improper. A sufficient answer to that may be that even in law long usage does not always conclusively establish propriety. Courts can hardly be said to have established an easement into the field of legislation, and the value of *stare decisis* is that the courts, even when wrong, are at least wrong predictably while here unpredictability is a primary feature of their behavior. We are searching for a model of proper judicial conduct, not trying to describe past conduct. To rest upon the observation that courts have always done the things objected to (though of course the cases discussed show that they have also refused to do them) is like telling a stockholder trying to introduce better cost accounting techniques that his firm has always failed to maximize profits.

[16] *Brown Shoe Co.* v. *United States,* 370 U.S. 294, 344 (1962).

In any event, many of the examples of contrary judicial behavior are explainable in other terms or according to other models. In constitutional law, for example, courts are sometimes forced to engage in unstructured political decisions precisely because they are dealing with values and principles which are by definition to be kept out of the hands of the legislature. Even in constitutional law, however, it is recognized as desirable that the Court, so far as the subject matter permits, should achieve the elements of judicial responsibility.[17] The common law is frequently cited as an accepted example of continuing judicial legislation. I do not think the common law today contains as much of the free balancing of contradictory values in it as people suppose. To the extent that it does it is very bad law. Much of modern common law, moreover, consists of rules built up by courts many years, even centuries, ago. The policy movements of today's common law courts are often merely interstitial, arising from the necessity to choose between conflicting values only in peripheral cases where it is not clear what rule governs. The major rules themselves were legislated by courts of a very different era, during a time when society was far less democratic and the legislature had not risen to its modern place as the direct representative of the people and the primary organ of policy making and political choice. The model of judicial behavior suitable to such an era is not appropriate to ours. A third example is the modern administrative process, but this does not provide a compelling analogy. It is common knowledge that the administrative process is performing poorly and one of the reasons is that legislatures and courts have not yet learned to make the agencies responsible by separating their executive, legislative, and adjudicative functions and requiring them to perform according to the models appropriate for each. Our present administrative process is thus not a model for emulation but a warning of what the judicial process may become if courts persist in mixing grossly legislative and adjudicative functions.

Courts do not appear to be comfortable making case-by-case political judgments. They are likely, if they adopt a theory that seems to require such *ad hoc* decision making, to seek an escape either by framing hard, arbitrary rules[18] or by placing the real decision-making function in some other agency. The latter technique may account for the rule in recent years that "the government always wins." The Supreme Court, encumbered by an incoherent economic theory of injury to competition through injury to competitors, the mystery of market definition, and the impossibility of reconciling rationally the contradictory values of consumer and small business welfare, may have turned

[17] See "Neal, Baker v. Carr: Politics In Search of Law," 1962 *Sup. Ct. Rev.*, 252.
[18] *Id.* at 300-327

the problem over to the Antitrust Division and may now be contenting itself with ratifying the decisions of the government.[19] This, of course, relieves the Court of operating with unadministrable criteria but it does not achieve judicial responsibility, for it creates a legislative agency where Congress intended none. The lawyers and economists of the Antitrust Division, moreover, are no more capable of solving the problems of antitrust policy than is the Court if they adhere to the Court's economic theory and its theory of the propriety of case-by-case political compromise between irreconcilable goals. Delegation of the problem to the Antitrust Division solves nothing. It merely puts the inherent irresponsibility of the decision-making process out of sight.

I have talked primarily about the impropriety of the goal of preserving small business under present statutes, but it should be clear that the introduction of any goal which conflicts with consumer welfare would be equally pernicious. I will briefly mention three which Professors Kaysen and Turner have classified:[20] (1) the attainment of desirable economic performance by individual firms and ultimately by the economy as a whole (by which they mean primarily economic efficiency and progressiveness); (2) the achievement and maintenance of competitive processes as an end in itself; and (3) a code of "fair" competition or conduct in the marketplace.

Each of these goals is partially or wholly inconsistent with the policy of maximizing consumer want satisfaction. The first goal stated is partially consistent since I take "efficiency" to be defined in terms of meeting consumer desires. I do not know how else one could measure the value of output. But the propriety of "progressiveness" as an antitrust criterion is not obvious. As defined by Kaysen and Turner, "progress consists in increasing output, in increasing output per unit of input by the development of new techniques, and in producing new and better final products."[21] Progress in this sense is obviously not costless to consumers. It requires the devotion of resources to research and development that would otherwise be devoted to the production of other goods and services. Progress will occur even without special consideration by the law, but the rate will be that which consumers choose by the degree to which they make it profitable to engage in the activity of producing progress. Courts have no criteria for establishing compromise deviations from consumer welfare here either. Such deviations should come from specific legislative direction, as in the patent statutes, subsidies, or tax relief.

The second possible goal listed—the achievement and maintenance

[19] See Simon, "A Partial Search for Affirmative Antitrust Answers," in *Basic Antitrust Questions in the Middle Sixties* (Nat. Indus. Conf. Bd., 1966), p. 31.
[20] Kaysen and Turner, *Antitrust Policy* (1959), pp. 11-18.
[21] *Id.* at p. 13

of competitive processes—presumably means something other than a consumer welfare standard which is already stated in the goal of efficiency. If it means preserving rivalry for its own sake, there seems no point in it. If it means maximizing competition, it does not state a goal that is even conceivable. A policy of maximizing competition would, as Justice Holmes pointed out in his Northern Securities dissent,[22] require the dissolution of virtually all industrial and commercial organizations. It is a prescription for the annihilation of our society and most of the individuals in it. Even a policy of pushing to a condition that a majority of economists would agree constituted pure competition would involve a vast destruction of the wealth of our society. And "workable competition" seems a meaningless concept unless it is merely another term for guessing about the impact upon consumers of moving from a given situation to an alternative. Then it becomes the consumer welfare standard.

The adoption of a code of "fair" competition, the third alternative, seems to mean the prohibition of certain commercial conduct on moral or ethical grounds. Antitrust, that is to say, could be in whole or in part a tort law. But the tort law concept of antitrust has serious difficulties built into it. When a tort rule prohibits a method of competing or creating efficiency, its effect is to favor some producers at the expense of other producers and at the expense of all consumers. That is unlike any tort law that I know of because the activity inhibited is entirely beneficial to the community. There is, in addition, the very real difficulty of defining tortious business behavior without reference to consumer welfare. The business community could not be expected to provide a useful consensus. Businessmen themselves have long been in disagreement, for example, as to whether vigorous price competition is "ethical."

The difficulty, in any event, of stating a general standard for tortious economic behavior, is illustrated by Kaysen and Turner's definition of economic "coercion" which might be the tort prohibited: "Typically, then, coercion consists in the ability of a firm with market power to impose terms in a bargain which the other party would refuse, were there an alternative transactor with whom he could deal more advantageously."[23] This definition of a tort standard demonstrates the inutility of such concepts. The requirement of "market power" either means that the tort exists only when present law would see illegal monopoly, in which case we are back to using a consumer welfare policy, or that any firm in a less than purely competitive market must somehow guess what terms and conditions it could get if the

[22] *Northern Securities Co.* v. *United States,* 193 U.S. 197, 400 (1904).
[23] Kaysen and Turner, *supra,* note 20, at p. 17.

market were purely competitive. If "market power" does not mean one of these things, then, under the definition given, every market transaction involves coercion—and by both parties simultaneously.

Nor is it any more helpful to employ other suggested standards, such as the equal treatment of parties in similar positions. If the parties are really equally situated, the market will treat them equally, for that will be the most profitable way to treat them. This truism can be avoided only by giving equality an ethical content which does not correspond to economic categories, and that raises once more the difficulty of defining "ethical" competition without reference to consumer welfare, and the question of why it is "ethical" for government to prefer one set of businessmen over another set and over the entire set of consumers.

These considerations lead me to conclude that the introduction of goals other than consumer welfare into antitrust is destructive of antitrust as law. Confining antitrust to consumer welfare, on the other hand, permits courts to employ the teachings of economic analysis to estimate whether the net effect of a particular structure, act, or agreement is likely to be an increase or a decrease in output. The judgment must often prove rough and may change over time as economic understanding progresses. But these kinds of uncertainty we must always live with in a legal system. One is uncertainty about where a spectrum will be cut. The other uncertainty arises from doubt as to the state of the courts' intellectual sophistication concerning the phenomena to be dealt with. Neither of these kinds of uncertainty can ever be eliminated in a legal system that undertakes to deal with matters of any complexity. The kind of uncertainty which a legal system ought not to tolerate, particularly where statutes are involved, is that which arises because judges are making case-by-case and ex post facto the political choices. We have in Congress a body of politicians elected to make political choices and to write them into prospective rules of general application.

Much more could be said on the topic of the goals of antitrust in a longer paper. But it seems to me both that the claims of the judicial process in this field have gone almost unnoticed and that they constitute the single most decisive argument in favor of confining antitrust to a consumer welfare goal.

VERTICAL MERGERS, MARKET POWERS, AND THE ANTITRUST LAWS*

By WILLIAM S. COMANOR
Harvard University

In recent years, there has been a rapid advance in the development of antimerger policy. Although the enforcement agencies frequently encounter adverse rulings in the lower courts, they have gathered a string of victories on appeals taken to the Supreme Court. This trend, in fact, has been sufficiently strong that a dissenting Justice could criticize his brothers by remarking that "the sole consistency that I can find [in the case at trial as compared with earlier ones] is that in litigation under Section 7, the Government always wins."[1]

Despite the flow of recent cases, there is still considerable dispute concerning some of the issues which underly this expanded enforcement effort. Of these, one of the most important refers to the circumstances in which vertical acquisitions have the effect of promoting and extending market power. In this paper, we shall examine this problem in light of the current status of enforcement towards vertical mergers.

The Extension of Market Position

Before proceeding, we should note that a firm's market power is conceptually different from its market position. It is defined in terms of behavior and refers instead to the ability to "behave persistently in a manner different from the behavior that a competitive market would enforce with a firm facing otherwise similar cost and demand conditions."[2] Although economists have noted that vertical integration by itself cannot create market power, they have argued that joining two successive stages of production can extend its scope and impact.[3] This is founded on the recognition that an integrated firm may have extensive leverage over nonintegrated rivals which are competitors at one stage of production and suppliers or customers at another. On this account, single-stage firms may be faced with the prospect of a price squeeze which at the least will absorb a large share of the rents associ-

[1] Dissenting opinion of Justice Potter Stewart in *United States* v. *Von's Grocery Company et al.*, 384 U.S. 301.

[2] Carl Kaysen and Donald F. Turner, *Antitrust Policy*, p. 75.

[3] See, for example, M. A. Adelman, "Integration and the Antitrust Laws," *Harvard Law Rev.*, 1949, p. 43.

ated with production and at the most will drive them from the market.[4]

The classic example of a price squeeze occurred in the aluminum industry during the period when Alcoa enjoyed a virtual monopoly in the production of the basic metal. To realize the gains from a discriminatory pricing structure, Alcoa found it profitable to set a relatively high price for aluminum ingot and a relatively low one for rolled aluminum sheet. The degree of discrimination was sufficiently great, moreover, so that for a number of years during the 1920's, the ingot-sheet price differential fell below the full cost of rolling sheet.[5]

The impact of this pricing structure on independent aluminum fabricators was clear. They were placed at the mercy of Alcoa, from whom they were forced to purchase the basic metal, and Alcoa held the power to extend its market position into the fabricating industries. Moreover, even if Alcoa's underlying intent was not to extend its monopoly position, the price squeeze was the natural result of a discriminatory pricing structure.

It is not surprising that the prewar aluminum industry furnishes the classic example of a price squeeze, for it also provides the classic case of monopoly. Alcoa enjoyed a high degree of market power and there was little fear of competitive pressure from existing firms or new entrants. Without this power, it is less likely that Alcoa would have pursued these pricing policies to the extent that it did for fear of driving independent fabricators into the arms of existing or potential rivals. Thus, it appears that the ability to extend a market position from one stage of production to another depends on the degree of market power enjoyed by the firm at the original stage. Moreover, this expansion may be viewed as a result rather than a cause of market power.

This process of extension is achieved through a form of foreclosure. Because of vertical integration, nonintegrated firms at the second stage are readily foreclosed from their suppliers or customers whenever an integrated firm moves to preempt a segment of the market. On this account, recent cases which concern vertical mergers and vertical integration have been argued and decided on the basis of a theory of foreclosure.

The Theory of Foreclosure

This approach to the problem of vertical acquisitions was enunciated clearly in the Brown Shoe decision of 1962. In deciding this case on vertical as well as horizontal grounds, the Chief Justice declared that

[4] This prospect is discussed in Corwin D. Edwards, "Vertical Integration and the Monopoly Problem," *J. of Marketing*, Apr., 1953, pp. 404-10.
[5] Donald H. Wallace, *Market Control in the Aluminum Industry*, pp. 379-95.

"the primary vice of a vertical merger . . . is that, by foreclosing the competitors of either party from a segment of the market otherwise open to them, the arrangement may act as a 'clog on competition,' which 'deprives rivals of a fair opportunity to compete.' Every extended vertical arrangement by its very nature, for at least a time, denies to competitors of the supplier the opportunity to compete for part or all the trade of the customer party to the vertical arrangement."[6]

In this approach, the degree of foreclosure is used to measure the extent of the market which is removed from competitors at both stages of production. If, for example, a firm with 50 percent of the market at an early stage acquires another who holds a 10 percent market share at the only succeeding stage, then as a result of the merger, rivals of the latter firm are said to be foreclosed potentially from half of the market for needed supplies, while rivals of the former are foreclosed from about 10 percent of the market for their output. In this manner, moreover, the theory is symmetrical and is used to refer to either stage of production.

Although these percentages measure the share of the market which may be preempted from nonintegrated rivals, they do not indicate even the prospective increases in market shares at the two levels. In the example above, the acquired firm at the latter stage may well be able to expand its market share to nearly 50 percent. Either through a price squeeze or a simple refusal to deal, second-stage rivals will be unable to obtain sufficient supplies at "reasonable" prices, and their market position is likely to deteriorate. At the same time, however, the acquiring firm at the early stage may find it difficult to expand its position very much simply as a result of the acquisition. Especially if it formerly had close dealings with the acquired firm, there may be few additional sales which result, and its market position may remain relatively stable.

In using the theory of foreclosure, its proponents assume that joining two successive stages of production will create a tried buyer-seller relationship which has the effect of removing these sales and purchases from the market. Not only does this affect the immediate market position of single-stage firms, but also it may force many of these firms to develop competing patterns of vertical integration as a defensive measure.[7]

[6] Opinion of Chief Justice Warren in *Brown Shoe Co., Inc.* v. *U.S.*, 370 U.S. 323, 324.
[7] In this context, the following statement appears in the annual report of a firm in the cement industry: "Vertical integration within our industry has been on the increase in recent years. *Alpha* is presently not inclined to integrate vertically. However, if our position in the industry is put in jeopardy as a result of such corporate arrangements, there will be no alternative but to make similar moves." *Annual Report of the Alpha Portland Cement Company*, 1963, p. 4, as quoted in the Staff Report to the Federal Trade Commission, *Economic Report on Mergers and Vertical Integration in the Cement Industry*, Apr., 1966, p. 3.

Although vertical integration may have a major impact on concentration at specific stages of production, as well as on the market position of nonintegrated firms, it does not necessarily follow that integration leads to an expansion of market power. Although concentration ratios tend to be correlated with market power, this finding does not imply that the sum of these ratios at succeeding stages of production is similarly correlated. The theory of foreclosure appears to be founded on market position rather than market power, and it is important, therefore, to examine further the effect specifically on the latter of vertical integration.

Integration and Price Behavior

It has been noted by others that vertically integrated monopolies may make only a single monopoly return.[8] In most circumstances, there is only a single final price at which profits are maximized throughout the various stages of production, and this price level will normally be reached both by a single-stage monopolist, who deals with competitive suppliers and customers, and by an integrated firm who holds a monopoly position at all vertical stages. The latter firm will set no higher price, for this would lead only to lower aggregate returns.

For given degrees of market power, vertical integration may in fact lead to lower prices than would otherwise result. In a situation, for example, where firms at an early stage of production have sufficient market power to set a price for their output which is below the monopoly price for the output of a succeeding stage by an amount equal to the costs of production—including the opportunity cost of capital—at the later stage, vertical integration will lead to an "improved" market result. In these circumstances, the final price established by a vertically integrated firm will be lower than that which would result from a series of single-stage monopolies. The same conclusion follows, moreover, if firms at a later stage of production have garnered sufficient market power to establish, on their own right, the joint-profit maximization price. Here, also, the fact of an internal source of supply will account for little as compared with an independent but purely competitive source. If, however, costs are higher due to market power at the earlier stage, then vertical integration will lead to lower prices and greater outputs than would otherwise exist.[9]

[8] See, for example, Robert Bork, "Vertical Integration and the Sherman Act: The Legal History of an Economic Misconception," *Univ. of Chicago Law Rev.*, 1954, p. 196.

[9] This position is similar to that taken in a number of earlier studies. See, for example, Professor Spengler's conclusion "that in an imperfectly competitive world vertical integration enables the higher-stage producer to evade 'monopolistic' surcharges imposed by suppliers in lower stages, thus putting him in a position where he finds it advantageous to ask lower prices than would be asked in the absence of vertical integration and in the presence of existing horizontal integration." Joseph J. Spengler, "Vertical Integration and Antitrust Policy," *J.P.E.*, Aug., 1950, p. 352.

A similar conclusion is obtained if elements of bilateral monopoly are present in markets between the two stages of production. In these circumstances, the output level at the early stage is lower than that which would be set by a single integrated firm so long as the bargain between buyer and seller is reached in terms of prices alone. Here, vertical integration will lead to greater purchases of intermediate products and, also, higher levels of final output. At the same time, however, if contracts are made which specify the quantities to be delivered as well as prices, then output levels in intermediate markets would equal those reached under vertical integration. And here, the achievement of integration would not lead to different market results.[10] Thus, vertical integration will foster greater output levels even if successive monopoly or oligopoly positions are not present so long as elements of bilateral monopoly exist in intermediate markets and if institutional factors prevent contracts founded on both prices and quantities.[11]

In light of these findings, it is instructive to examine one claim which has been made regarding the anticompetitive consequences of vertical integration. The claim is that leading firms in specific manufacturing industries have extended their market power by acquiring control of scarce raw materials so that no new competitor could present himself.[12] In these circumstances, however, a scarce resource limits the degree of competition at the mining rather than the manufacturing stage of production, for it is at this stage essentially that entry restrictions are imposed. On these grounds, the high returns earned by vertically integrated firms are likely to consist primarily of the rents associated with a scarce resource rather than of gains arising directly from vertical integration.

To recognize this conclusion, we need to consider what difference it would make if manufacturing firms in this market situation were forced to divest their mining affiliates. In the absence of horizontal divestiture, prices at the extractive stages would still exceed costs, and vertical divestiture is likely to lead to higher rather than lower prices for output at the manufacturing stage. Even if high concentration levels in mining are achieved through the process of vertical integration, because opportunities are first recognized by firms in related manufac-

[10] The analytics of this argument appear in George J. Stigler, *The Theory of Price* (rev. ed.), pp. 240-41.

[11] Professors Machlup and Taber maintain that in situations of bilateral oligopoly, contracts are more likely to refer to prices alone rather than to both prices and quantities, and that therefore "the conclusion that vertical integration might lead to an increase in output could still be maintained." Fritz Machlup and Martha Taber, "Bilateral Monopoly, Successive Monopoly, and Vertical Integration," *Economica*, May, 1960, p. 112.

[12] For example, it is argued that "the maintenance of the monopoly position of Alcoa, in the period after the expiration of its basic patent, was aided greatly by its purchases of bauxite mining companies in Europe and its policy of acquiring leases on power sites in advance of their development and use." Kaysen and Turner, *op. cit.*, p. 121.

turing industries, the high returns which result are founded essentially on market power garnered at the constituent horizontal stages of production.

Integration and Market Structure

The previous analysis concerns the effect of vertical integration on pricing practices which are determined for given degrees of market power. It is founded, however, on the assumption that the degree of market power is independent of the extent of vertical integration. Although it presents some important conclusions, we need to examine also the effects of vertical integration on specific elements of market structure, and on the process through which market power is achieved. And it is here where the major impact on competition is likely to exist.

In earlier studies, attention has focused largely on market concentration. Although vertical integration may well have the important effect of extending high concentration levels from one stage of production to another, it cannot be held responsible for high concentration prior to integration. And it is the latter that results in market power which is exercised through vertical relationships. In addition, since vertical integration is relatively unlikely to provide a vehicle for the extension of market shares where existing shares are greatest, integration should not lead to higher concentration at the stage where market power is founded. Thus, the competitive impact of vertical integration is probably best examined in terms of its influence on remaining structural and behavioral elements.

Entry barriers are a significant feature of market structure, and it has often been suggested that the degree of vertical integration has a considerable impact in this area. Among the most significant of these barriers are those which are due to important economies of scale, from absolute cost disadvantages associated most frequently with a scarce resource and from high absolute capital requirements which impede entry as a result of imperfections in the capital markets.[13] As we noted above, however, entry barriers which result from a scarce resource will exist regardless of the extent of vertical integration. And also it appears that although vertically integrated firms may extend their market position into succeeding stages, they are unlikely to extend similarly their aggregate returns.

A similar conclusion appears in the case of significant scale economies. It has been argued that if scale economies are important at one stage of production but inconsequential at another, vertical integration will lead to higher overall entry barriers and thereby to greater market power. The problem with this argument, however, is that it assumes

[13] See Joe S. Bain, *Barriers to New Competition.*

implicitly that the degree of market power achieved through entry barriers is additive at successive stages, and this is precisely what the analysis above shows to be incorrect. Although vertical integration may, under these hypothesized circumstances, restrict entry into the second stage, new firms will be confronted with no greater barriers, on this account, to entry at both stages than they did, prior to integration, at the first stage alone. And, therefore, integration should not lead to increased returns. This conclusion, we should note, follows from the fact that vertical integration has not had a direct impact on the height of entry barriers as they exist at individual stages of production.

The effect of vertical integration on the entry barriers associated with high absolute capital requirements is, however, likely to be more significant. To the extent that vertical integration inhibits entry at a single stage of production, a new entrant is compelled to begin operations at both stages. And this must necessarily raise the amount of capital required for entry.[14] We should note that entry barriers heightened in this fashion result precisely from the joining to two stages of production, and thus, vertical integration may well contribute to the achievement of enhanced market power.

This result, however, requires the additional condition that firms limit their supply and purchasing relationships to internal sources and refrain from dealing with external suppliers and customers. Only in this manner will it be made clear to prospective entrants that they must begin operations at both stages of production, and that they must accept thereby the costs and risks of entry at a larger aggregate size. Thus, these barriers appear to depend also on the reactions of integrated firms to new entrants for they exist only if single-stage entry is effectively barred.

In situations where new entrants can readily achieve product "quality" and costs which are no more desirable than those of existing firms, then the latter are likely to adopt policies which discourage the encroachment of new entrants. For this reason, vertically integrated firms are unlikely to accede to the development of the stable and satisfactory commercial relationships which may be necessary for the success of a new venture. Potential entrants will be forced, at the very least, to recognize that existing, integrated firms have effective leverage in price setting through the ease with which they can shift their

---

[14] Professor Bork concedes that capital requirements would be greater if vertical integration fostered conditions under which new firms would be forced to enter at both stages. But, he argues, this should not restrict entry "unless there are impediments in the capital market that prevent capital from flowing to areas where it can most profitably be employed." (Bork, op. cit., p. 195.) Professor Bain finds, however, that high capital requirements do have the effect of imposing significant entry barriers in a number of industries. Bain, op. cit., pp. 156-66.

dealings to internal facilities. And it is unlikely, therefore, that new single-stage firms will be able to garner even a small share of the economic rents which are normally associated with production. We can note, moreover, that it is in this context that the theory of foreclosure, including the prospect of a price squeeze, finally provides some analytical assistance. As compared with their vertically integrated rivals, single-stage firms face a highly precarious existence which should serve to impede the entry of new firms at a single stage.

Even in circumstances where the product technology and cost functions of the new entrant are more advanced then those of existing firms, single-stage entry will still be effectively restricted. Despite the fact that lower costs or higher revenues in the short run would result from dealings with single-stage competitors, existing integrated firms will face the prospect that short-run maximizing behavior is likely to lead to more competitive market conditions and a long-run decline in monopoly profits. To the firm, entry barriers are an asset which has value similar to that of a new machine or a well-received trademark, and, therefore, firms can be expected to adopt policies which can be explained only as an "investment in entry barriers."[15] And here, the short-run costs associated with restricting sales and purchases from competing single-stage firms may comprise a portion of the costs of this investment. In this fashion, there may be a conflict between long-run behavior designed to promote entry barriers and short-run profit maximizing behavior. Thus, so long as the short-run costs are not excessive, single-stage entry is likely to be difficult regardless of the prospective products which might be offered and the costs which might be achieved by new firms.

Before concluding this discussion, we should consider whether vertical integration necessarily implies a reserved buyer-seller relationship regardless of the prices offered by competing firms. This question, however, has no general answer, but rather depends on the calculus of investment decision making. Its answer turns on the costs associated with a refusal to deal with independent suppliers and purchasers, and on the benefits, in terms of maintained or expanded market power, which might result from requiring new firms to enter at both stages of production.[16] Although vertical integration does not imply reserved commercial relationships in all circumstances, this trading pattern

[15] This phrase was used originally by Thomas A. Wilson in discussion at Queen Lake, Mass.

[16] In the automobile industry, and specifically with regard to General Motors, it appears that this "investment" has not been made. Rather than limit its dealings to affiliated suppliers, General Motors instructs its purchasing agents to buy from independent firms if they can obtain their requirements at a lower price than that set by a G.M. subsidiary Peter F. Drucker, *Concept of the Corporation,* p. 50. Given the overwhelming entry barriers which already exist in this industry, this behavior is not surprising.

should exist whenever it is likely to lead to higher restrictions on entry and a greater degree of market power.

A further element of market structure is the scope of product differentiation, and it is in this area also that the competitive effects of vertical integration appear to be significant. As is well known, product differentiation plays a dual role. Not only does it influence the nature of competition among established firms, but also it is likely to have a major impact on conditions of entry. For this reason, it is important to note that integration between manufacturing and distribution stages of production serves frequently to enhance the extent of product differentiation and to limit correspondingly the degree of price competition. Although this prospect may have major competitive significance and deserves more analysis and elaboration, we shall here do little more than repeat Professor Bain's conclusion that the "integration of retail dealer-service organizations by manufacturers, either through ownership or exclusive dealing arrangements," serves frequently as an important source of the entry barriers attributable to product differentiation.[17] Since these barriers, where they exist, result precisely from joining two successive stages of production, anticompetitive consequences can be attributed directly to vertical integration.

A Structure-Behavioral Interaction

Vertical integration may, in some cases, contribute to the achievement of market power in a further manner. This concerns the prospect that structural and behavioral factors may interact to open a route through which vertically integrated firms may bypass markets which resist oligopolistic control in favor of others which do not. In the achievement of market power, behavioral factors are significant, and these include the ease by which reaction patterns among rival firms can be established which embody a recognition of mutual interdependence. Since these behavior patterns may vary widely among markets at succeeding stages of production, an important incentive may exist for vertical integration so that a firm's output can be sold in a more rather than a less congenial market situation.

As we noted above, vertical integration may enable firms to extend their market position from one level to another. To the extent that the latter market is more susceptible than the former to the development of oligopolistic pricing policies, this extension will lead to enhanced market power. This result, however, requires that firms at the original stage of production have sufficient power to extend their market position into the second stage, while at the same time, not be able to gain all of the advantages of market power. Although this prospect clearly

---

[17] Bain, *op. cit.*, p. 142.

depends on some peculiar market circumstances, this interaction may be significant in some instances, and in fact appears to contribute, in some measure, to the achievement of market power in the primary copper industry. For this reason, it is instructive to outline the characteristics of this industry which lead to this result.

Although the copper industry is highly concentrated in the smelting and refining of primary metal, it is far less so when secondary metal is included within industry boundaries.[18] When this correction is made, the four-firm concentration ratio drops from 90 percent to 64 percent.[19] In addition, the leading four firms in this industry are all vertically integrated into metal fabricating and control approximately 62 percent of total fabricating capacity.[20] Although copper fabricating is carried on in four distinguishable industries, concentration ratios at approximately this level are probably reached in the two industries of the four which consume over 80 percent of refined copper output.[21]

Although the markets for fabricated copper products appear conducive to the development of normal patterns of oligopolistic rationality, this is not the case in the market for refined copper. This difference exists because of the peculiar cost characteristics of secondary producers. Rather than being a manufacturing firm which sets production levels through a calculus of established costs and demands, secondary producers resemble purchasing agents whose output is controlled by available supplies. When these are abundant, their cost curve may be quite low although when supplies are limited, costs become prohibitively high. As a result, their output levels depend heavily on factors which are largely exogenous to the industry and thereby not liable to effective oligopolistic control. Furthermore, supplies of scrap metal frequently come to the market at a very uneven pace, and this fact accentuates the inability of the leading firms to control the prices of refined metal.

Not only does this process create price instability, but also it has the effect of impeding the development of accepted behavior patterns in price setting. In attempting to stabilize prices, the dominant firms often follow rather than lead price changes. Even more important, frequent price changes create problems of oligopolistic communication and emphasize divergent interests among the leading firms. In many mar-

---

[18] Various observers have noted the high degree of substitutability which exists between primary and secondary copper. See James L. McCarthy, "The American Copper Industry, 1947-1955," *Yale Econ. Essays*, Spring, 1964, p. 81; Frederick T. Moore, "Industry Organization in Non-ferrous Metals" (unpublished Ph.D. dissertation, Univ. of California, 1951, p. 120); and A. D. H. Kaplan, *Big Enterprise in a Competitive System* (rev. ed.), p. 157.

[19] U.S. Senate, Subcommittee on Antitrust and Monopoly, *Concentration Ratios in Manufacturing Industry—1958*, 87th Cong., Second Sess., pp. 141, 142, 166.

[20] McCarthy, *op. cit.*, p. 73.

[21] These four industries and their share of total copper consumption are: Brass Mills, 45 percent; Wire Mills, 37 percent; Foundries, 16 percent; and Powder Plants, 1 percent. *American Metal Market*, Sept. 27, 1965, Sec. 2, p. 27.

kets, conflict can be resolved by the expedient of doing nothing, a solution of the *status quo ante* is not unusual, but when external factors are important, this solution may not be possible. And conflict leads directly to more competitive results.

In these circumstances, vertical integration appears as an important expedient which contributes to the achievement of market power. Because of fabricating requirements for a steady supply of metal, the leading copper refiners were able in the past to extend their market positions despite their inability to control price levels for refined copper. Through the ownership, moreover, of fabricating facilities which assume a dominant position within the relevant markets, these firms have frequently achieved high profits and market power despite low prices for refined metal.

Conclusions

From the analysis above, the theory of foreclosure does not appear as a useful tool with which to examine the relationships between vertical integration and market power. Although this theory may well describe the process by which market shares are extended from one stage of production to another, it does not describe a comparable extension of market power. To the extent, moreover, that antitrust policy is directed toward limiting market power rather than merely market shares, this theory is an inappropriate means of determining the competitive consequences of vertical integration.

At the same time, however, vertical integration is clearly a significant factor which contributes to the achievement of market power in a number of situations, and this result has important implications for antitrust enforcement. Not only may it have a substantial impact on the height of entry barriers, by requiring that successful entry take place at two successive stages of production, but also integration between manufacturer and distributor is likely to be a major element in the creation of product differentiation. In addition, there may be some circumstances in which structural and behavioral factors interact, as they do in the primary copper industry, to limit the degree of competition. Thus, it appears that while vertical integration may have important competitive consequences, these can best be demonstrated through their effect directly on market structure and behavior rather than through a theory of foreclosure.

Although the competitive impact of vertical mergers may be appraised by considering the effects of the vertical integration thus created, a further route lies in viewing customers and suppliers as likely potential entrants into related stages of production and examining whether a particular merger will remove an important potential en-

trant. In this manner, vertical acquisitions may be considered on the same basis as other mergers between firms which are not direct competitors. An earlier study, moreover, has concluded that "the most important advantage of a prohibition of certain kinds of vertical mergers lies, it seems, in the temptation it creates for the monopolists or monopsonists on each separate stage to integrate forward or backward by establishing new facilities. Whereas vertical integration by merger may block the increase of competition in the industry, competition may be effectively increased when firms extend their facilities to enter production at the stages of their erstwhile suppliers or customers."[22] But whether we appraise the competitive consequences of vertical acquisitions through their removal of potential entrants or through the direct effects of vertical integration, we need to distinguish between market position and market power, and to make these appraisals explicitly in terms of the latter.

[22] Machlup and Taber, *op. cit.*, p. 117.

CONSCIOUS PARALLELISM AND THE
KINKED OLIGOPOLY DEMAND CURVE

By WILLIAM HAMBURGER
University of North Carolina

I. *Introduction*

Oligopoly theory deals with the behavior of small groups of firms whose members operate interdependently. It is a topic that embraces a diversity of often incompatible theories. These range from the extreme of individualistic decision making (as represented by the theories of Cournot and Bertrand) to explicitly collusive behavior. Between these extremes are the possibilities of partially collusive behavior. In such cases a different mechanism is employed and intermediate consequences in terms of production and pricing policies are attained.

These intermediate cases have received much attention because they seem particularly pertinent in a world which limits the possibilities of rigorous collusion, yet contains many industries in which individualistic decision making seems inappropriate to the decision-makers. Sweezy's "kinked demand curve" and Chamberlain's "mutual dependence recognized" belong to this category. I shall discuss an adaptation of these theories, which is related to Boulding's concept of "perfect oligopoly," and which I shall call "conscious parallelism" to suggest something collusive, but not altogether collusive. In addition to stating a theoretical description for this case, I shall attempt to analyze its implications with respect to the consequent level of prices and with respect to the flexibility of prices.

II. *Decision Making by Extremism*

Let us assume that an oligopolistic industry's behavior has somehow been determined up to a single magnitude; for instance, the industry-wide level of prices. This sweeping assumption provides a basis for the theories of Chamberlain and Sweezy, referred to above, and I doubt if any other interpretation of these theories is feasible.

If this assumption or precondition is accepted, we may view the determination of the magnitude in question as a particularly simple problem in social decision making, to be dealt with by such methods as dictatorship, democracy, or extremism. The term "dictatorship" suggests some form of coercive price leadership. "Democracy," in the context of oligopoly price determination, equates with collusion. "Extremism" is a particularly interesting method, precisely because it does not require

direct recourse to either of these other practices, so that it may be usable when the others are impractical or illegal.

Under extremism, the individual desiring a greater (or in an alternative case, a lesser) decision magnitude than his fellow participants sets the pace. A game may, for instance, end when the first participant tires, or a war may last as long as the most patient participant wants it to last. In the case of oligopoly price determination, the industry price may be established at the level which is preferred by the firm which wants a lower price to prevail than the prices its competitors desire.

If we accept this theory of decision making by extremism for a particular group of firms, we can expect to observe aspects of price leadership. However, the leader would not necessarily be the largest firm nor the lowest-cost firm nor even the same firm for successive changes in industry price. It would merely be that firm which favored the lowest price for any reason whatever, including ignorance of its own best interests, as well as such rational considerations as lower costs or competitive pecularities vis-à-vis its fellows or with respect to firms outside the industry in question.

If several firms desire the prevailing price, each of them, as well as all followers, will have a demand curve that is kinked at this price. It would lower the industry price by lowering its own price, but it would merely remove itself from the group of equal-price firms were it to break the rules of the game by establishing and persisting in a higher price. But if the prevailing price were the exact price desired by only one of the firms, while higher prices were preferred by all others, this firm would be a true leader, since others would, within limits, follow its price increases as well as its price reductions. This leading firm would, however, encounter "resistance" (namely, the kink in its demand curve) were it to attempt to raise the price above the level that is desired by the next potential leader among the remaining firms. Its "kink," in other words, would occur at the next highest desired price rather than at the prevailing price which it has established.

III. Conscious Parallelism

I have discussed "decision making by extremism" in a context of a single decision that can be considered without reference to other decisions. The assumption that this procedure is a valid one, for a given industry, together with the assumption that the decision is made by extremism, may be called the assumption of "conscious parallelism"; where several magnitudes are pertinent, it requires that they be bound together by natural or institutional constraints, so that the profit position of each firm is fully determined by the magnitude that is determined by the extremist mode of decision making. Let us say, for in-

stance, that price and quantity are pertinent magnitudes, and that prices are determined by extremism. The quantities may be determined directly by the prices, either as a result of imperfect substitutability of the products in question or through market-sharing practices —either case would constitute conscious parallelism, by this definition, and would be distinguished from a situation in which the quantities were determined, not only by the prices, but also by the identity of the price leader.

IV. *Implications of Conscious Parallelism*

The mode of behavior described above would lead to consequences identical with those of perfect collusion if all firms were alike. However, firms are not, in general, identical in their cost structure or market position, and even when they are, they may differ in their calculation of what is best for themselves. The asymmetric structure of extremist decision making suggests that differences among firms, whether real or fancied, will not "average out," but will display themselves in the market behavior of the group of firms.

The larger such a group becomes, the greater the diversity of wishes that may be expected to prevail among them. This means, for instance, that for a collection of plants with given cost conditions, the price will be lower if a smaller number of firms controls them. Whether this provides for a Cournot-type transition between monopoly and competitive price cannot be determined without reference to the conditions which determine the degree of diversity that is compatible with conscious parallelism. A high degree of diversity could conceivably lead to subcompetitive prices, if it did not first destroy the postulated industry structure.

Another consequence to be expected when firms have diverse interests is an asymmetry in the industry's price flexibility. While each firm is capable of leading the industry to lower prices, when either its own circumstances or those of the industry change in a fashion that makes this action profitable to itself, only the leading firm can expect to be followed in price increases. Assuming that there is some cost involved in unsuccessful attempts at price changes, any uncertainty as to the identity of the leader would lead to hesitancy by all firms to respond to, say, an increase in costs. Consequently, we may expect that prices under conscious parallelism would be more inflexible in the upward direction than in the downward direction, distinguishing this mode of oligopoly behavior from both flexible prices and from the symmetrically inflexible prices that are associated with other modes or theories of oligopoly behavior.

DISCUSSION

JOHN S. McGEE: My remarks are confined to Comanor's paper. As I read it, vertical integration may increase or prolong monopoly by raising entry barriers, by increasing product differentiation, or by permitting oligopolists to sell only in those markets that are easiest to control. I will examine these indictments in turn.

Comanor's work shares the present general confusion about what entry "barriers" are and how they work. It is now common to include as barriers any costs or requirements of entry, however irreducible or natural they may be. Thus, if secretaries had to read and write, that would "bar" illiterates from secretarial posts; flat-footed, obese intellectuals may already be "barred" from professional football; and so on.

In any case, vertical integration allegedly increases the capital required for entry since, to avoid "squeezes" and other perils, any entrant must begin as fully integrated as his rivals. And—as we have long been told—the more capital required, the higher the entry barrier tends to be. Although some of the flaws in this argument are discussed in the very sources he cites, Comanor plunges on. I waive the important questions of what a squeeze really means in this context, and whether it is more probable or powerful than straightforward price-cutting of the kind in which even nonintegrated firms could engage. Nor will I dwell on the degree of monopoly necessary to force an entrant to integrate. Even so, there remain the problems of cost savings, necessity for multiple-level entry, and capital requirements and monopoly rates of return.

Perhaps Comanor's greatest novelty is his neglect of efficiency as a motive for vertical integration. In his world, cost reductions result only from reducing monopoly markups. This not only ignores much integration actually observed. It also leads to bias: an integrated firm may buy from outsiders to minimize cost; but Comanor takes this behavior as evidence that "other" barriers to entry are "overwhelming." If integrated firms do not buy from outsiders, that is evidence that they are making an "investment in entry barriers."

If integration lowers costs, entrants will be forced to integrate, as they should be. If costs are not reduced by integration, as Comanor apparently assumes, it is not clear why any single entrant has to integrate. More or less contemporaneous entry by different firms, each operating on one level, would efficiently satisfy their mutual needs.

But probably the best way to get at the capital requirements bogey is to play Comanor's own biased game. If entrants conclude they must be vertically integrated, they will require more capital. The question is whether this increased capital is an entry barrier.

Monopoly returns are defined as the excess over what is necessary to induce the relevant resources to work. The integrated firms are assumed to be earning them. Normal returns are just adequate to attract the necessary factors, including risk-bearing, capital, etc. Higher risk, just as any other higher cost, calls for higher returns even under competition. But, no matter how costly or

risky an industry is, if outsiders expect sufficiently high returns, they will want to enter. How high expected returns must be to exceed "normal" will thus depend upon the variance and upon the degree of risk-aversion.

Assertedly, larger amounts of capital are impossible, more difficult, or more expensive to obtain. As Bain once put it, they may "not be forthcoming from savings out of salary or from the winnings in a poker game." True enough. Nevertheless, there is a capital market, and abnormally high returns await. If the entrant is competent to enter and the capital market is not hopelessly flawed, he should be able to borrow or obtain equity capital. If no single investor is willing to gamble so much in a single venture, the capital market can do what it does every day: find several or many, giving each a share. Rounding up the requisite number for an absolutely larger investment may be more expensive; but monopoly rates of return are higher than necessary to induce capital and the other factors to do their work.

Besides, existing firms, which are at least as large as the would-be entrant, are earning rents. How did they get the capital? It may be argued that they did it the easier and cheaper way: out of earnings. That argument may imply that monopoly returns have been calculated erroneously, using irrelevant cost functions. Or existing firms may have rare talents or extremely efficient internal capital "markets," whereas no firm using external capital markets can now do the job as cheaply. Or perhaps the shortage is of people who can run large firms. But all of this begins to sound like monopoly based on superior skills— a most undemocratic notion; or that capital markets do not work, which has yet to be demonstrated.

Of course, even a finding that absolutely large firms have their own superior internal capital markets generates no expectation of widespread monopoly. For there are lots of absolutely large firms, and the theory never required that entry occur only through newly created firms.

Following Bain, Comanor also convicts integration of increasing "product differentiation," thus reducing competition. It is unclear whether this is due to increased promotional efficiency or to something else. Casual consideration of the degree of vertical integration and advertising intensity in soap, cigarettes, razor blades, and so on, makes one wonder. Similarly, the empirical work of Telser, Stigler, and Pashigian, for example, lends little or no support to the notion.

My last concern is with Comanor's "Structure-Behavioral Interaction," which he illustrates with a two-page study of the copper industry. The "model" is so underspecified and the facts so fragmentary that it is hard to tell what is going on. Apparently, though, it is that the demand for primary copper fluctuates sharply, partly because of fluctuations in the supply of secondary copper, whereas fabricated copper markets are much more stable. Because oligopolists do better when life is not too complicated, fabricated copper turns out to be more tractable. I am not sure what this has to do with vertical integration. Nor am I convinced that Comanor has dealt suitably with several questions, including market definition and concentration, the influence of vertical integration upon costs, the rationality of achieving "stability" by ignoring changes in the opportunity costs of inputs that are supplied internally, and

the relative merits of stabilizing prices in the face of fluctuating costs or stabilizing prices in the face of fluctuating demands.

PETER PASHIGIAN: On discussing William Hamburger's paper, I feel somewhat like the captain of a ship approaching an iceberg. The captain sees only the icecap. I see, but only faintly, the elements of a theory of oligopoly. The captain can only conjecture how large the submerged portion of the iceberg is. I can only conjecture how these elements will be ultimately packaged and whether a theory of oligopoly with testable implications will emerge.

How are these elements packaged at present? Hamburger begins by assuming that all firms will not desire the same industry price. Hence, there will be a distribution of desired prices. Each firm prefers its price to be the industry price. How this distribution of prices is derived is not completely spelled out. At one point, Hamburger suggests that these prices may be derived from subjective or irrational considerations. I believe this approach is unpromising. Under another interpretation we may assume that each firm calculates the price that maximizes its profits under the assumption that other firms will match this price. For simplicity, we may assume that each firm's market share is determined by nonprice variables. If significant differences in demand and cost conditions exist, these conditional desired prices will differ.

Presumably, low-cost producers will tend to desire lower industry prices than high-cost producers. We may rank these desired prices from lowest to highest. The industry price will be determined by the firm with the lowest profit maximizing price. I believe this is what Hamburger calls decision making by extremism although he may not agree with the method of determining the lowest price. As stated, there is nothing in the specification that says that the largest firm will quote the lowest price, i.e., will be the price leader, or that the same firm will quote the lowest price over the cycle, since relative cost and market positions among firms will normally change over the cycle. Thus, the identity of the price leader can change more frequently under this specification than under the "price leadership" model.

If there is a difference between the lowest profit maximizing price and the second lowest price, the demand curve facing the price leader will have a kink at the second lowest price because all other firms are willing to follow a price increase to the second lowest price. All other firms will have a kink at the lowest price since a price increase will not be followed by the price leader.

What implications can be drawn from this hypothesis of price determination under extremism? First, the level of price depends on the diversity of these conditional desired prices. Hamburger conjectures that the larger the number of firms in an industry, the greater the diversity in cost or revenue conditions and hence the lower the level of price.

A second implication of price making by extremism is the upward inflexibility of prices. Suppose there is a cost to an unsuccessful attempt at a price change; e.g., a temporary or permanent reduction in market share. The identity of the price leader may change if there is an increase in the price of a factor. If the former price leader raises price, the sales of the price leader and his followers will decline, at least temporarily, if a new price leader emerges.

Apparently, the cost of an unsuccessful attempt to raise price produces the rigidity. The uncertainty of the identity of the leader is irrelevant if there is no cost to an unsuccessful price change. I am not sure why the same conditions would not lead to a downward rigidity as well. If a cost reduction leads the price leader to reduce price, he must fear the emergence of a new price leader and a consequent loss of sales if a still lower price is quoted. Before a definite conclusion can be reached, a more detailed specification of the determinants of each firm's sales is required. Nevertheless, one test of Hamburger's proposition is to determine if prices are more flexible in industries with a recognized and persistent price leader than in industries where the role of the price leader changes hands.

These are the elements of price determination by extremism. Let me list several possible limitations of this approach.

In a number of instances Hamburger refers to differences between firm cost curves to explain the distribution of conditional desired prices. It seems preferable to rely on differences in market position rather than difference in cost levels to explain the distribution of desired prices. If there are differences between the marginal costs between firms, there is an incentive to merge. We would expect these cost minimizing forces to work toward the gradual elimination of differences in marginal costs between firms.

Hamburger's approach to the oligopoly problem differs from two recent contributions to the oligopoly problem. In Stigler's paper on oligopoly and the paper by MacAvoy and Orr, there is an attempt to determine the conditions under which price cutting by a rival leads to a large or a small gain in sales or profits of the rival. Thus, Stigler is concerned with the effect of the number of producers, the number of buyers per seller and the probability of repeat sales on the expected percentage increase in sales by a potential price cutter. The cost and method of detecting a price cut is the principal concern. I do not detect a similar concern in Hamburger's paper. Once the industry price has been set by the extreme member, followers may still have an incentive to cut price given that the original price leader will not be able to detect a price cut instantaneously. To assume that price changes can be detected and matched instantaneously, robs the analysis of much realism.

Another point of concern is the attempt to fashion a theory which generates implications about price rigidity. Recent studies of list and transaction prices suggest that the apparent rigidity of wholesale prices is in large part an illusion. The studies by McAllister, Fleck, and Stigler indicate that transaction prices are considerably more flexible than are the prices reported by the Bureau of Labor Statistics. The results of these preliminary studies of transaction prices suggest that theories which attempt to explain rigid prices should have a lower popularity rating among economists. Certainly, the expected rate of return from additional investment in this area has been reduced as a result of these studies.

ECONOMIC HISTORY

JOHN LAW OF LAURISTON:
BANKER, GAMESTER, MERCHANT, CHIEF?

By EARL J. HAMILTON

University of Chicago and
State University of New York at Binghamton

John Law of Lauriston may not have been at one time the richest and most powerful uncrowned person that Europe had ever known, as he claimed after his fall; but in neither respect was his claim absurd. At the peak of his power and wealth, early in 1720, he dominated the Royal Bank, or first Bank of France, as no other man has ever dominated a national bank, not excepting Hjalmar Schacht or Montagu Norman. He was in complete control of the Company of the Indies which had a monopoly on the importation and sale of tobacco; the privilege of operating the twenty-six mints; a monopoly on trade with Africa, Asia, and the East Indies; the farm of the indirect taxes; the entire national debt; a monopoly on the fur trade of Canada; and possession of Louisiana which then comprised more than half (and economically a very good half) of what is now the United States, excluding Alaska. He owned the Hotel Nevers (which now houses the Bibliothèque Nationale), the Mazarin Palace, more than a dozen impressive country estates, several plantations in Louisiana, a high percentage of the real property around the present stock exchange, more than a third of the magnificent buildings that now adorn the Place Vendôme, and at least a hundred million livres tournois' worth of stock in the Company of the Indies. He was Contrôleur Général des Finances, or Finance Minister, to whom the provincial intendants (who, Law maintained, were the real rulers of France)[1] were responsible. Above all, he had the ear, the confidence, and the affection of the Duke of Orleans, a Regent with almost absolute power. Despite strong and numerous enemies, indeed John Law was a chief.

Consisting largely of fancy real estate and of stock in a company feverishly sought by the well-to-do throughout western Europe, Law's affluence aroused curiosity, and long before he became Finance Minister it was well known that he had enormous influence at court. Law's

---

[1] As of 1718, the Duke of Saint-Simon, a member of the Council of the Regency, said that "Bâville [has been] for thirty years king and tyrant of this great province [Languedoc] under the name of intendant." *Mémoires de Saint-Simon*, Boislisle ed., Vol. XXXIII (Paris, 1922), p. 11.

handsome features, great histrionic talent, bold ideas, and magnetic personality were hardly less striking than his wealth and power. His having been a refugee from justice under sentence of death for killing a notorious spendthrift in a duel enhanced his notoriety. The thunderous collapse of his System was no less spectacular than its rise, and the cruel losses it inflicted upon millions who had fancied themselves rich also fastened attention on Law.

The fact that Lady Catherine Knowles, daughter of the Earl of Banbury, mothered Law's children and passed as his wife while married to another man made Law and Lady Catherine reticent and secretive about their lives. At the height of his glory Law was so busy that it was next to impossible to see him, and after the collapse of his System humiliation, persecution by creditors, and fear of insult or attack by men ruined in his Mississippi affair kept Law to himself.

Keen interest in Law has induced an immense literature on his life and work; and lacking facts, writers have shrouded them in myths and legends. We are told *inter alia* (1) that he was a professional gambler, (2) that by gambling he met and won the friendship of the Duke of Orleans, whose support was vital to all his operations in France, (3) that he promoted lotteries, arranged the odds in his favor, and profited handsomely from them, and (4) that he had had no experience in business, banking, or finance when he began his operations in France.

That John Law of Lauriston was a professional gambler is one of his easiest characterizations to document to the satisfaction of scholars. Many writers on Law have told us that he was expelled from Paris in 1707 because he was too successful at the games of chance he had introduced.[2] In a letter to Nicholas Desmarets, Director General of Finance, concerning his passport to leave France, then required of foreigners as well as of Frenchmen, Law stated as a matter of fact that he was being expelled because he had proposed to give France paper money superior to gold or silver.[3] It stands to reason that Desmarets knew why Law was being driven out and that a misrepresentation would have been ruinous to a projector who firmly intended to show up again.

Much more important is the fact that on November 16, 1708, the Marquis of Torcy, who as Foreign Minister of Louis XIV had to be an expert on intelligence, wrote the Marquis of Argenson, one of the shrewdest and most efficient police commissioners Paris has ever had, that "Law a Scot whom you knew well when he was here last year and whose profession is to be a gambler [*et dont la profession est d'être*

---

[2] For example, Émile Levasseur, *Recherches Historiques sur le Système de Law* (Paris, 1854), p. 17.

[3] *Archives Nationales* (Paris)—hereafter AN, G⁷ 12.

joueur] has returned without a passport. One assures me that his intentions are not good and that he is serving our enemies as a spy."[4] Torcy requested a full report on Law's activities and intentions. Since he could not be found, the Marquis of Argenson did not reply until July 22, 1714, when he informed Torcy that "a Scot named Law, gambler by profession and suspected of evil intentions toward the King appears at Paris in high style and has even bought an impressive home in the Place Louis le Grand [Vendôme], although no one knows of any resource [he has] except fortune in gambling, which is his whole profession. I cannot believe that the motives which have aroused just suspicions against him have ended with the peace." But in his own hand Torcy wrote on this letter: "He is not suspect. One can leave him in peace."[5] So the just suspicions must have been political in nature and not caused by a belief that he was living by gambling, in which Torcy concurred. Indeed he did not need to be, for he had enough invested in French rentes to support him in high style, and in partnership with the Earl of Islay he was holding and actively dealing in British securities at London on a large scale.[6]

In an important sense, however, one of Law's most fatal mistakes was a gamble—for both political and economic effect. In the late summer of 1719 the outlook for stock in England seemed bright enough to tempt not only English but some continental speculators to switch from shares in Law's Company of the Indies to English stock. Maddened by jealousy and fearing a flight of capital, Law began to belittle English securities and to boast that he not only could but might break the market. His remarks were reported to the British foreign office and caused a minor alarm.[7] To prove his disdain of English stock, Law sold Thomas Pitt, Earl of Londonderry and uncle of William Pitt, £100,000 of East India stock, one of the bluest of all English blue chips, short for £180,000 for delivery by August 25, 1720.[8] The price of East India stock not only withstood the shock but rose so high that George Middleton, Law's London banker, who supplied the margin to protect Londonderry, was strained to the utmost to raise it.[9] From April, 1720, to the delivery date, when Middleton bought the final £50,000 of shares at 315 to 335, his chief efforts were to supply the margin and to buy stock to cover this

[4] *Archives des Affaires Étrangères*—hereafter AAE, Mémoires et Documents: France, 1161, fol. 128.
[5] AAE, Correspondance Politique: Angleterre 262, fols. 286-87.
[6] Archives of Coutts and Company, Ledgers for 1712-1714. Cf. Historical Manuscripts Commission, Report on the Manuscripts of His Grace the Duke of Portland, preserved at Welbeck Abbey, John Drummond to Earl of Oxford, Utrecht, May 8, 1713, Vol. V, 1899, p. 287.
[7] See, for example, Public Record Office, Pultney to [Craggs], Paris, January 6, 1720 N.S., SP 78/166. Cf. Coutts and Company, Letter Book 014, p. 343.
[8] Coutts and Company, Second Show Case
[9] Coutts and Company, Letter Book 014, pp. 213-15.

short sale with minimal effect on the quotation.[10] Desperate appeals to Law for assistance were sent through his brother William and George Skene, one of his closest business associates;[11] and the imperious need for remittances to Middleton was a major depressant of the external value of French currency.[12] Owing to Middleton's commitments and Law's inadequate assistance, Middleton, one of the greatest bankers of his era, failed on December 13, 1720 O.S., or on Christmas eve, N.S.— a sad time to fail.[13] Law lost on this one ill-fated short sale about $970,000, approximately one-fifth more than the $800,000 John Jacob Astor wrote Albert Gallatin, on October 9, 1815, while trying to persuade him to go into business with him as a partner, and hence not inclined to underestimate his fortune, that he had accumulated and invested.[14] This was after Astor had worked hard and saved hard in the United States for almost thirty-one years. This bear operation was a significant factor in the collapse of Law's System, and the staggering debt resulting was a sore in his flesh to the end of his days.[15]

Very little is known about Law's life from the spring of 1692, when it is believed he went to London upon obtaining his majority, and the spring of 1701, when he was imprisoned at Paris, except that he quickly ran into debt and lost his patrimonial estate of Lauriston, which was repurchased by his mother and later entailed to him. This suggests the possibility of gambling losses. But it seems more likely that extravagant living and losses in speculative business were responsible. From 1701 to the end of Law's life he was far too busy writing treatises on money and banking, formulating projects for national banks, striving desperately to persuade rulers to charter a bank or a trading company, trading in goods and securities to support his family and earn a fortune, administering his own gigantic and national affairs in France, attempting to salvage something from his estate, subtly trying to persuade the Regent to recall him to France, and defending his ideas and policies to have time for gambling either professionally or habitually.

A widely believed myth is that Law had ingratiated himself with the Duke of Orleans by gambling with him on earlier sojourns in Paris, was his crony when he settled there in December, 1713, and had easy access to him when he had tracts or projects to submit to the crown. Once when Louis XIV desperately needed funds he showed plutocratic Samuel Bernard his gardens at Marly, to the horror of the Duke of

[10] Coutts and Company, Letter Book 014, pp. 193, 213-14, 343.
[11] See, for example, Coutts and Company, Letter Book 014, pp. 471, 477, 498, 526.
[12] Cf. Coutts and Company, Letter Book 014, p. 315.
[13] Coutts and Company, Letter Book 014, pp. 508-09.
[14] Kenneth W. Porter, *John Jacob Astor: Businessman* (Cambridge, Mass., 1931), Vol. II, p. 596.
[15] Bibliothèque Méjanes, *John Law to William Law* (London), Apr. 2, 1722, Ms. 614, fols. 204-05.

Saint-Simon;[16] but, though Bernard's fondness for gambling was well known, he was not invited to the nightly gaming in the apartment of MMe. de Maintenon.[17] Neither by gambling nor in any other way had Law won the friendship of the Duke of Orleans before he became Regent. In fact, Law had to go through devious channels and to implore the assistance of obscure and humble people to gain access to the Duke of Orleans in the last years of Louis XIV. It was by extraordinary knowledge and personal charm in serious discussions of monetary theory, banking institutions and public finance, not by gaming, that Law won the friendship and admiration of the Duke of Savoy and the Duke of Orleans, as well as of the dissolute and corrupt Duke of Bourbon.

Histories of gambling differ sharply from biographies of Law regarding him as a gambler. Writers on Law who classify him as a gambler, as most do, say that he played at dice, cards, faro, chemin de fer, and the like and patronized such gambling parlors as Duclos, Poisson, and Gesvres.[18] Most histories of gambling say nothing of Law in their discussions of gamesters. The few that do mention him support their indictment by saying he was responsible for the Mississippi Bubble, with its frenzied speculation in the Rue Quinquempoix, Place Vendôme, and the garden of the Hotel de Soissons, without mentioning cards, dice, other games of chance or gambling dives.[19]

Further evidence that Law was not a professional gambler is the fact that George Middleton, who knew him well, was not only his London banker and agent but banked, financed tobacco imports, and supervised construction of ships in British yards for his Company of the Indies. Though Law was deeply in debt to him from the collapse of the Mississippi Bubble in 1720 to his death in 1729, Middleton served and aided him with unwavering patience and understanding. Since Middleton consistently placed a high estimate on sobriety and dependability, it hardly seems possible that he would have become or remained intimate with a professional or compulsive gambler.[20]

The temptation to disclose that Law had been a professional gambler at a definite time and place and to give particulars would have been too great for his enemies or even his associates to resist either when he was the idol of Europe or when he was a pathetic object of sympathy after his fall, if he had had such a past. The lack of such an indictment strongly supports other evidence that Law never made a profession of gambling.

[16] *Mémoires*, Vol. XVI (Paris, 1925), pp. 34-37.
[17] Simonde de Sismondi, *Histoire des Français,* Vol. XXVII (Paris, 1842), p. 197.
[18] See for example, Barthélemi Marmont du Hautchamp, *Histoire du Système des Finances* (The Hague, 1739), Vol. II, p. 1; Émile Levasseur, *op. cit.,* pp. 16-17.
[19] See, for example, Jean Dusaulx, *De la Passion du Jeu* (Paris, 1779), Pt. I, pp. 81-83.
[20] That one could not trust a gambler as a business agent was stressed by Jacques Savary in his *Le Parfait Négociant* (8th ed., Paris, 1721), Vol. I, Pt. II, pp. 254 ff.

Nevertheless, Law did gamble occasionally, as did almost all the elite in his day. For example, in mid-March, 1719, he bet the Duke of Bourbon a thousand new louis d'or that there would be no more ice that winter or spring. Just before April seventh "the Duke sent him two cart loads of ice from his basins at Chantilly. Mr. Law had already sent him the thousand louis d'or."[21] But Law may have made this bet to let the Duke win for political reasons. On a few occasions he had gambling losses in both Italy and England, after the Mississippi Bubble, but none were serious.[22] Once Law sought advice while toying with the idea of taking a proffered wager of ten to one that he would not be recalled to France, adding that if the stakes were high enough to pay his debts, he would be sorely tempted.

Law wagered ten thousand to one at times and twenty thousand to one at others that one could not throw a designated number with six dice at one throw. Since the odds against this were six to the sixth power, or 46,656, to one, Law fared well if he had many takers. He also offered to pay anyone a thousand pistoles for throwing six sixes at one throw if his opponent would pay him two pistoles every time he threw four or five sixes. According to the celebrated mathematician Nicolaas Struyck, the odds were 4,870 to one thousand,[23] or nearly five to one, in Law's favor. Since the cost per venture was low and the potential prize high, this play should have appealed to gamesters; and it offered Law an opportunity to make money, to appear bold to gamblers and observers by hazarding a large sum, and (by winning) to magnify his image as a calculator, which he fervently desired. There is no evidence that Law achieved spectacular financial success through these schemes. Recreation and fame seem to have been his chief motives.

Law acquired a reputation for projecting and establishing lotteries in Holland, using the methods of Thomas Neale in England, with the chances supposedly calculated in some mysterious fashion so as to give him an advantage; and on this account he is said to have been driven out of Holland.[24] I have been unable to find in the Royal Archives (*Algemeen Rijksarchief*) or the Municipal Archives (*Gemeente Archief*) of Amsterdam any project for a lottery presented by Law. If a plan drafted by him had been accepted, it probably, but not necessarily, would have been preserved; and owing to the excellent organization of the Dutch archives, the chances are good that it would have been found.

[21] *Archivio di Stato*—hereafter AS, Torino, Philippo Donaudi to Marchese di Borgo, Paris, Apr. 7, 1719, Lettere Ministri: Francia, Mazzo 147.

[22] See, for example, Bibliothèque Méjanes, John Law to Marquis de Lassay, Venice, July 11, 1721, Ms. 614, fol. 86.

[23] *Uytreekening der Kanssen in het Speelen* (Amsterdam, 1716), p. 15.

[24] Mr. Gray, *The Life and Character of the Great Mr. Law at Paris* (London, 1721), pp. 15-16.

Law's known connection with the Dutch lottery was the opposite of the legendary one. He devised a plan for insuring holders of national lottery tickets offered for sale beginning on August 12, 1712, against drawing blanks. For about half a year beginning in the early fall of 1712 he advertised in the *Haarlemse Courant, Haagenaar,* and other newspapers of Holland that anyone holding tickets in this lottery might register 10 with him at his home in The Hague or with his agent, Isaac Abrabanel, in Amsterdam, pay a fee of 100 guldens and receive 300 guldens if all 10 were blanks or register 15 tickets for the same fee and receive 500 guldens if all drew blanks. As time progressed, Law varied the scheme by insuring at a very low rate if all prizes above a certain large figure would go to him. Nicolaas Struyck calculated the odds on Law's original proposition and found that if the law of large numbers operated, he made a gross profit of 22 guldens 12 stuivers out of a hundred,[25] or 22.6 percent. The deposits of Law and Abrabanel in the Bank of Amsterdam suggest that large numbers did come into play but that the business did not flourish.[26]

Again Law used his knowledge of probability in an effort to make a gain on a game of chance, as does everyone providing insurance. It seems that Law originated this scheme, for I have been unable to find advertisements of other insurance against losses in the Dutch lottery in any newspaper from six months before until six months after his advertisements ran. So the story that by some form of intervention Law made a languishing Dutch lottery work must be doubted.[27] Otherwise, either Law would have continued his operation; or in a country with such abundant capital and bold entrepreneurs someone would have followed his example. Law's scheme did provide holders of lottery tickets an opportunity to reduce their chance of loss by reversing their position, but this meant that a ticket holder made two bets with odds on each against him. Offered by an insurer lacking Law's glamor, this insurance might have found few buyers among cautious and prudent Dutchmen.

Vittorio Amadeo II of Savoy, staggering under the burden of restoring Turin after its long siege and the heavy start-up costs in governing Sicily (added to his dominion by the Treaty of Utrecht), turned his attention to lotteries in the summer of 1715. On October 29 Philippo Donaudi, secretary of his embassy in Paris, followed instructions to send the Duke all the legislation establishing lotteries, tontines, and life annuities in France since 1689 and reported that he had conveyed

[25] *Op. cit.,* p. 90.
[26] *Gemeente Archief,* Amsterdam, Amsterdamse Wisselbank, Rekening Courant, Registers No. 140-No. 144.
[27] Historical Manuscripts Commission, Portland MSS., John Drummond to (Earl of Oxford), Utrecht, May 8, 1713 N.S., Vol. V, p. 287

to Law the Duke's request to send him plans for a lottery. Law sent him statutes authorizing lotteries in England and Holland and copies of the structure of certain existing lotteries made by his young son. He was too busy preparing for the extraordinary session of the Council of Finance, on October 24, to consider his plan for a royal bank, to reply immediately.

In a letter of December 7, 1715, Law wrote the Duke that he was supplying plans for a lottery loan, apparently what the Duke had in mind, but urged him not to establish one. "Even supposing that it would be fully subscribed, I am of the opinion that Your Majesty would do better by borrowing the sums you need in the ordinary manner at 5 or 5.56 percent (*denier vingt ou denier dix-huit*)[28] than to set up lotteries. Public lotteries are less bad than private ones, but they are injurious to a state. They do harm to the people, take the paltry sums they earn by their labor, make them dissatisfied with their lot, and give them a desire to grow rich by gambling (*hasard*) and luck. Servants lacking money are tempted to steal from their masters to obtain means to play in the lottery. And the bourgeois use in lotteries the money which should sustain their trade and pay their creditors. These kinds of projects should not be permitted in well governed states." Law admitted that lottery loans might attract money from abroad but argued that retaliation would draw money from Savoy. Law also advised against borrowing on life annuities. He would have to pay 10 percent, and with the rate of interest at 5 percent, an annuity "on the life of a healthy man is worth about 8 percent." He sent plans for a lottery loan that would cost the Duke 6½ percent for prizes and interest; but doubting that funds could be obtained at this rate, he sent an alternative plan that would cost 7½ percent.[29] It is curious that Law, who was unusually consistent for his time, stated that investors overestimated the combined return from prizes and interest on lottery loans, yet argued that funds could be obtained more cheaply by rentes, where the return was clearly understood.

In Law's projecting years lotteries were flourishing in Italy, France, and Holland; and a big majority of his numerous bank-projecting rivals provided for lotteries to raise either a part or all of the capital required. Law's projects never did, and neither did he resort to a lottery in the summer or fall of 1720 when he was searching desperately for

---

[28] In December, 1715, the interest rate on long-term loans was 5 percent at Paris and Toulouse. The long-term rate at Amsterdam was 4⅓ percent, and the rate on loans of twelve months or less was 4 percent. Oddly, the one call-loan rate I have for Amsterdam was 6 percent. Sources, all contemporaneous manuscripts, will be given in my book on "John Law's System," in preparation.

[29] AS, Torino, John Law to Duke of Savoy, Paris, Dec. 7, 1715, Sez. 2ᵃ, Finanza, Mazzo 1.

means to deflate the redundant French currency and the stock of the Company of the Indies.

Law's detractors have argued that he had had no banking or business experience before he founded the Banque Générale. Actually he had grown wealthy in business at Genoa; and, as we have seen, he had had limited experience in insurance and had not only invested heavily in government rentes in Paris, but had dealt in securities on a large scale in London. Notorial records show that borrowing and lending money had been a way of life with Law. He had dealt in foreign exchange and had kept accounts for years in the Bank of Saint George and the Bank of Amsterdam. While "exiled" in England from October, 1721, to August, 1725, Law turned to importing and selling paintings and tapestry. For paintings George I was one of his best customers.

While operating in France, Law showed little inclination to take a risk in competitive enterprise. Such monopolistic privileges as exclusive issue of notes payable to bearer on demand and requirements that royal revenues be held in bank notes were quickly secured for his Banque Générale. His Company of the West and his Company of the Indies were shielded by monopoly in virtually everything they undertook. His own funds were invested largely in stock in the Company of the West and Company of the Indies, the price of which he attempted to peg, and in real estate, which he wanted in order to prove his permanent commitment to France and because he thought it was safe. He also invested considerable sums in government annuities for his family. He did take a risk on plantations in Louisiana, as did almost every leading banker and businessman in Paris, including the severest critics of his System and his worst enemies. And he took a big risk when he failed to make investments abroad as a hedge against loss of favor in France, but this may have been because he thought the political danger of such a course would be still greater.

Law's investment decisions suggest timidity; and so do his studious efforts to prove he was cool, courageous, and bold by taking chances, though mathematically improbable ones, to lose a large sum in gambling with dice on infrequent occasions when a large gallery was looking on.

In many respects the political power Law had under an absolute monarch was not good for him. It led him to rely upon force, at the sacrifice of principles, rather than voluntary action in self-interest. For example, he repeatedly stated in early tracts and banking projects that forced circulation would depress rather than raise the value of paper money. Yet he resorted not only to forced circulation but to despotic action against specie, including search and seizure, to combat deprecia-

tion of his bank notes. Though he had argued that legal regulation could not reduce interest rates, he set a legal maximum rate and bitterly fought the Parlement, or appellate court, to sustain it. He reacted furiously to criticism in English and Dutch newspapers and exerted pressure through diplomatic channels to have them censored, only to be reminded by foreign ministers and ambassadors that in those countries the press was free. Against violation of his tobacco monopoly and fraud in excise taxes he used strong-arm methods; and, though his recruitment for Louisiana was much more humane than historians believe, impressment, transfers to seaports in chains, and imprisonment until embarkation hindered emigration to Louisiana, which was dear to his heart.

GOVERNMENT REGULATION AND GROWTH IN
THE FRENCH PAPER INDUSTRY DURING
THE EIGHTEENTH CENTURY*

By WARREN J. SCOVILLE
University of California, Los Angeles

Papermaking apparently originated in China about two centuries before the Christian Era. Arabs learned the art in the eighth century and diffused it along North Africa, whence it spread into Spain in the twelfth century, into Italy in the thirteenth, and into France either near the beginning or near the middle of the fourteenth century. The industry prospered in France, and by the middle of the seventeenth century numerous mills established on streams around Angoulême and in Auvergne were supplying not only French markets but were also exporting large quantities of fine writing and printing paper to England, Holland, and nearly every other country in Europe. Around 1700 the French economy was in the middle of a protracted period of stagnation, and the paper industry was especially depressed. Several intendants stated at that time that papermaking had greatly declined over the past two decades, and the newly established Council of Trade learned from the survey it ordered in 1701 that the number of active mills was everywhere much smaller than formerly.

A number of factors had been responsible for this decline. In northeastern France the War of the League of Augsburg had both destroyed a great many mills and closed the foreign markets of those that managed to survive. Louis XIV's wars also provided the infant paper industry in both England and Holland with almost complete protection from French producers. Mills in these two countries rapidly expanded their output and permanently captured most of their domestic market from the French. The French government at about the same time unfortunately increased its taxes on paper because of military expenditures and also extended its regulation and control over the industry with the ostensible purpose of protecting the quality of its output. In addition, the revocation of the Edict of Nantes in 1685 helped to reduce France's comparative advantage in papermaking by forcing many of her skilled workmen to flee to England, Holland, and Germany, where they diffused many of the cherished secrets and skills which had

made French paper the most popular in Europe.[1] Hence warfare, increased taxation and control, and the diffusion of technical knowledge and skills by refugee Protestants during the last decades of Louis XIV's reign brought an end to the industry's prosperity in the seventeenth century. How the state both hindered and aided its subsequent revival forms the subject of the present inquiry.

Anyone who has worked in French economic history, especially prior to 1800, must have been vexed by the paucity and unreliability of almost any sort of quantitative data. I hesitate to assert that reliable figures do not exist for limning the industry's growth; instead, I shall simply state that I have not succeeded in uncovering them. I hasten to confess also that I do not think economic-historical investigations should necessarily be confined to those problems and to those areas and eras where quantifiable (and reliable) data exist for testing hypotheses. Some problems may deserve our attention and some tentative conclusions or hypotheses may prove valuable even though they rely upon inferences drawn from incomplete, questionably accurate, and even qualitative data.

During the last third of the eighteenth century more than 600 papermills were probably at work in 26 of the 36 *généralités* into which France was divided for fiscal and administrative purposes. This may have been twice the number at the opening of the century, when the industry was especially depressed, but was not much more than the total in the 1730's, after recovery had set in. Although most mills operated with only one vat, a great many at the end of the century had two and several had five or more. Nicolas Desmarest, an inspector of manufactures and an authority on papermaking, estimated that there may have been at least 900 vats in operation around 1788; and a survey the government conducted just before the Revolution suggests that the total may have been 1,050. I infer from such evidence that the output of paper in 1789-90 was probably three, and perhaps four, times the output for 1700-01. Since the annual output at the beginning of the century was exceptionally small, this comparison exaggerates the secular growth. Were the yearly output before the Revolution compared with the average around 1680 or 1720, which were more representative years, the increase would probably amount to only 100 percent.

The French industry expanded appreciably less than the British. Professor Coleman, in his excellent study of papermaking in Great Britain, has estimated, from far more reliable evidence than I have found for France, that the British industry quadrupled in size.[2] The

[1] Warren C. Scoville, *The Persecution of Huguenots and French Economic Development, 1680-1720* (Univ. of California Press, 1960), pp. 230-37, 327-29, 346, 440.

[2] D. C. Coleman, *The British Paper Industry, 1495-1860: A Study in Industrial Growth* (Oxford: Clarendon Press, 1958), pp. 89-90.

difference in these growth rates reflects the fact that in the seventeenth century British papermaking was an infant industry (with perhaps 100 mills in the 1690's) and had a small output, whereas the French was well developed. Also, English mills could easily expand as soon as wars and protective duties assured them a larger and larger portion of the domestic market. On the other hand, as the French saw their traditional English and Dutch markets slip away, they had to seek out and develop new ones, a task far more difficult than that faced by the English and Dutch, who simply expanded to fill a void in existing markets.

Like their counterparts in England, most papermakers in France worked only one vat. This meant that they operated on a small scale and employed on the average about three or four skilled artisans, one or two apprentices who assisted them, and from three to five women and children who sorted, cleaned, and cut the rags or who sorted and bundled the finished product. The master usually worked alongside his workers. Most mills hence employed from eight to ten people to make anywhere from 2 to 10 reams of paper a day or from 600 to 3,000 reams a year, depending upon the kind and quality of the output. The annual outputs most frequently given by government officials ranged from 2,000 to 3,000 reams (which represented from ten to twenty-five tons in weight and from 4,000 $l.$ to 9,000 $l.$ in value) if the mills worked regularly; if, as frequently happened, a shortage of water caused by freezing temperatures in winter and by dry spells in summer restricted production to six or eight months, the annual output was even less.

The representative mill with a single vat did not enlarge its staff during the century and hence did not significantly expand its output, for there were no technological innovations which increased worker productivity. The beating of rags into fibers constituted the first stage in production. Women first cut up and sorted the rags, allowed them to soften or rot for several weeks in moistened piles, and then transferred them to the pulping mill, which usually consisted of six to ten mortars, each equipped with stampers which were activated by water power. Around the middle of the century the Hollander or Dutch cylinder, which triturated the rags between its blades and a fixed metal plate somewhat as an old fashioned lawnmower cuts grass, slowly began to replace the stampers. It, too, was hydraulic. The second stage began with the transference of the fibers to the working vat, where an occasional stirring kept them suspended in tepid water. A skilled artisan then gathered enough fibers to form a sheet of paper on a wooden and mesh mold which he dipped into the vat and withdrew in such a way as to spread and entangle the fibers. After most of the excess water had drained off, another workman inverted the mold so that the matted

fibers fell onto a felt pad. When a sufficient stack of these had accumulated, they were squeezed several times in a large press to extract the remaining water and were then carried to the drying loft and finishing room. There less skilled workers completed the third and final phase of papermaking. They hung the sheets to dry, then dipped those which were to serve for writing into sizing and redried them, ironed and smoothed the sheets with blocks of stone or wood, and finally sorted and wrapped the finished product. It should be noted from this simplified version of the technology of making paper in the eighteenth century that only the pulping process was mechanized and utilized water power. The technological breakthrough in forming and finishing paper occurred during the early years of the following century, when an English manufacturer successfully developed a continuous paper-forming machine which a Frenchman named Nicolas-Louis Robert had conceived and patented in 1799.

Most papermills in the eighteenth century, some of which were converted oil or flour mills, represented relatively small capital investments, ranging from 4,000 *l.* to 20,000 *l.* Even so, many operators could not afford to own them and had to lease them from merchants or local landlords. Frequently, such an owner contracted with his lessee to buy and market the mill's entire output at some stipulated price and to supply all the rags and other materials needed in production. Important paper merchants sometimes built and leased several mills to different operators; more rarely, some enterprising papermaker leased and operated more than one mill.

While the typical French papermill still possessed only one vat and operated on a relatively small scale throughout the eighteenth century, manufacturers built more and more mills which, with their two, three, or four vats, resembled small capitalistic enterprises. Each of these employed from 25 to 100 persons of varying degrees of skill and produced two or three times as much output as the average. Because of the larger amount of capital invested in plant and equipment, they were often owned and operated by partnerships (*sociétés*), even though the members may have been of the same family. The industry also came to draw more of its capital and entrepreneurial talent from the outside.

Beaumarchais, author of *The Barber of Seville,* bought the Arches and Archettes mills in Lorraine in 1779 and sold them for 50,000 *l.* nine years later. These works together had five vats and employed around a hundred men, women, and children. Two famous families of papermakers who I understand are still in the business—the Montgolfier and the Johannot—each owned and operated extensive mills of from five to nine vats at Annonay in Southern France, employed con-

siderably more than a hundred persons (one document credits Pierre Montgolfier with providing employment for 300), made up to 400 tons of paper a year, and maintained storehouses in Paris, Orléans, Lyon, and at least three other places. The Dupuy family owned and operated the Grandrive factory, which the intendant described in 1767 as "the most considerable and most useful in all Auvergne." From 1747 through 1780 its annual shipments averaged almost 100 tons and were valued at 83,500 *l.* It had four vats and around a hundred employees. Nicolas Kranz and Joseph Villiet each employed a hundred or more persons in the two or three mills they each owned and operated in the Vosges. A paper merchant named J. B. Reveillon, who painted paper for covering walls and lining boxes, acquired and successfully operated a mill at Courtalin in Brie after 1770. When he liquidated his company in 1791, its assets were valued at more than 1,000,000 *l.* At about the same time at least two important factories were constructed in Provence with four and eight vats, respectively; the intendant of Angoumois wrote in 1778 that a large mill had recently been completed at Montbron, which represented an investment of 300,000 *l.;* an association of entrepreneurs built another with twelve vats at Sisteron in the Lower Alps in 1785; and the model factory of Langlée near Montargis in Orléanais had plans for operating twelve Hollanders to supply pulp for its thirty vats. Altogether there may have been a score of enterprises making paper on a relatively large scale near the end of the century. They helped lift the average number of vats per mill from around 1.2 in the early 1700's to something like 1.6 or 1.7 before the Revolution. It is interesting to note that Professor Coleman has concluded that the average in England and Scotland rose in similar fashion from 1.2 in 1738 to 1.5 in 1805.[3]

Everyone who has studied French history during the seventeenth and eighteenth centuries is familiar to some extent with the state's pervasive yet ambivalent role in the economy. Much of this stemmed from Colbert, who thought that the state should not only encourage but should also actively guide the economic and industrial growth of the country. He often extolled the merits of competition and economic freedom; yet at the same time he was convinced that the government should be prepared to institute direct controls under certain circumstances and during the early stages of new developments. The ministers and lesser officials who came after him do not appear to have had the same genuine interest in stimulating economic development. They also lacked his imagination and adaptability, and they acted as though more government regulations and controls were always preferable to less. After 1749, however, Vincent de Gournay, Charles and Claude

[3] *Ibid.,* p. 150.

Trudaine, and Turgot eventually reversed the trend toward more and more government control either by repealing previous regulations or by refusing to enforce them.

Many economic historians suggest that this liberalization of post-Colbertian statism induced a marked expansion in most industries. Papermaking was no exception. While at no time during the century did the government fail to interfere with the industry in one way or another, its actions on the whole after 1750 were less restrictive of growth and, especially, of entrepreneurial initiative than earlier. And most of the expansion in the industry which I have described and, especially, the increase in the number of capitalistic enterprises operating on a relatively large scale materialized during the second half of the century.

Colbert had first seriously attempted to regulate papermaking in 1671, without, however, fixing the size and weight for each kind of paper. Even though his successors reaffirmed the order in 1688, no one seriously attempted to enforce it for very long. A regulatory code appeared for mills in Auvergne at the end of 1727; a similar one appeared in 1730 for mills around Limoges and Angoulême; and a third modified the Auvergne code two years later. The most comprehensive one was first promulgated in 1739 and then modified slightly in 1741. Its sixty-one articles regulated practically every aspect of the industry for the entire country. It forbad mills to produce paper of different size and weight than was contained in a detailed schedule, unless the entrepreneur first obtained permission to do so from the intendant in the area. This meant, for example, that a mill was not free to fill an order for some foreign customer who wanted a special kind of paper. To assure that all paper conformed to the prescribed dimensions, old molds had to be replaced with new ones, which first had to be inspected by appropriate authorities and then branded for later identification. Manufacturers had to construct sheds over rotting bins and stamping mills; they had to build reservoirs and filters for the water used in washing fibers; they could not use lime or similar corrosives to whiten rags or hasten their disintegration; and no one could employ any machine other than stampers in the macerating process without explicit government permission. The law also regulated the hiring and training of apprentices, the hiring and dismissal of journeymen, the registration of "master fabricants," and the hours during the day when employees could make paper. Manufacturers in each district were to meet annually and elect two or four guards who were supposed to visit each mill from time to time and to see that the regulations were being followed.

Even though government officials consistently failed to enforce all its provisions and ceased to pay it any attention after about 1760, the

code nonetheless technically remained the law until the Revolution and did considerable mischief for a decade or two after its adoption. Such requirements as discarding old paper molds and having new ones made and inspected, or building sheds over the stamping mills, occasioned mills considerable expense. The attempt to standardize output discouraged experimentation and dissuaded entrepreneurs from producing what the market demanded. Nicolas Desmarest, who probably did more than any other single individual to diffuse technical information among papermakers during the last quarter of the century, urged the government to repeal the law. "In reading these regulations and comparing their provisions with the actual state of papermaking in France," he wrote, "one easily realizes how dangerous it is to try to control an industry by permitting only certain approved processes, which have to be changed and improved each day as new knowledge and needs develop."[4] According to him, the prohibition to use lime for decomposing rags and whitening pulp had been most unfortunate, and the sanction against macerating machines other than stampers retarded the diffusion of the Hollander.

This device offered many advantages. Developed by Dutch papermakers around 1670, it required less power than stampers and hence could be activated by windmills as well as by weak water mills. Furthermore, manufacturers no longer had to water-soften their rags, a process which lasted from one and a half to three months; cylinders reduced the time required for pulping by two-thirds; they required less upkeep, less supervision, and less operating space than stampers; and they produced a cleaner and more uniformly pulped batch. On the other hand, they were harder to install, represented a somewhat larger initial capital investment, and antagonized paper workers who may have foreseen that cylinders would cause mills to expand their plant and output. There is also some evidence that cylinders chewed the fibers into shorter segments than stampers and hence made a somewhat more fragile product. This technical disadvantage, however, was not significant except in the production of a few grades of paper for which durability was essential.

Despite its net superiority, Frenchmen adopted the Hollander slowly and with hesitation. Its eventual diffusion was due to the actions of certain entrepreneurs as well as to a more enlightened attitude of government. In 1737 two men discussed before the Royal Academy of Science whether manufacturers should attempt to substitute cylinders for stampers. This discussion may well have been the reason the government included the prohibition of cylinders in its code two years

[4] "Art de fabriquer le papier," *Encyclopédie Méthodique: Arts et Métiers Mécaniques* (Paris: Panckoucke, 1788), V, 587-88.

later. An entrepreneur at Montargis in Orléanais, nevertheless, obtained governmental dispensation in 1741 to install a cylinder on an experimental basis, as did four other mills during the next two decades. An order dated May 10, 1763, finally rescinded the 1739 sanction and allowed paper manufacturers to use whatever machines they wished. Largely through the efforts of Desmarest, who had just been appointed inspector of manufactures at Limoges and through those of a resourceful mill manager and engineer named Ecrevisse, who helped several factories install the machine, the number of cylinders increased rapidly. Altogether I have been able to identify at least fifty-nine mills throughout France which used the machine before the Revolution.

While the state finally eased its regulation of the production process and thus stimulated the industry's progress, its tax policy partially neutralized these liberal efforts. Ostensibly to insure that paper measured up to certain standards of quality but in reality to secure more revenue for itself, the government at various times (1633-48, 1656-1720, 1748, and 1771 to the Revolution) required all paper made in the realm to be inspected, marked, and taxed. Certain kinds used for business, legal, or official purposes (*papier timbré*) had to acquit special taxes after 1673. Paper, like most other commodities, had to pay tolls, local taxes, and regional customs as it moved over rivers or roads on its way to market. The restrictive effect which all these charges (amounting roughly to 20 percent) had upon domestic production was offset to some extent by the removal of all export taxes on paper in 1701-02 and by the collection of protective duties on all foreign imports.

In the field of labor relations the government vainly attempted to enforce rules which, on the whole, favored employers over employees. Papermakers were not organized into formal guilds, and the great majority of employers were skilled workmen who had managed to buy or lease some existing mill. They worked alongside their employees and shared their lodging and meals. Instead of the friendly family relationship which one might expect to have existed, there was everywhere considerable friction between masters and workers. The latter belonged to powerful, yet informal and illegal, associations (*compagnonnages*) which effectively set working rules and intimidated employers with threats of strikes, blacklisting, fines, and physical violence. The state vainly tried to change this. Numerous orders declared all worker associations to be illegal; an individual could not leave his job without having first notified his employer six weeks in advance; they could not prevent a master from having as many apprentices and journeymen as he wished nor from hiring papermakers from outside the immediate area; and they could not start work before three o'clock in the morn-

ing nor produce more than half their daily output before noon. The purpose of this last restriction was to increase discipline and saner living by preventing workers from completing their daily stint as early in the morning as possible and from carousing the rest of the day.

While these regulations would have strengthened the industry's competitive position had they been enforced, another set of rules offset this by limiting the freedom of employers. As a consequence, employers could not hire individuals away from other mills by offering more than the prevailing wage; they could not employ anyone who did not have a written discharge from his previous employer; they could not train boys as apprentices whose fathers followed other crafts unless there were no sons of papermakers available; they could not discharge a worker without first giving him six weeks' notice; and they had to feed their staff on holidays as well as on work days. Local officials, at least sporadically, tried to enforce these rules governing employers and employees. They fined employers for illegally recruiting workmen from other mills and disciplined employees for walking out on their work or for creating disturbances. The state, however, was unable to curb the power of the workers' associations or to protect sufficiently the property and decision making of entrepreneurs. The failure of the government in this area and the continued dominance of paper workers over their employers appreciably restrained the industry's expansion.

Another limiting factor was the cost of linen rags, which accounted for about two-thirds the annual operating expenses (including all wages) of a one-vat mill. And here the government acted more wisely and effectively than in the matter of labor. The supply of rags at any one time was relatively inelastic as to price, and the increase in supply over time was largely a function of the increase in population and the relative popularity of linen over other kinds of fabrics. Although the price quotations I have seen for rags are scattered, they leave the impression that this material increased in relative value during the century. At least, paper manufacturers thought so, for they continually complained about the relative scarcity and high price of rags. From 1697 to 1733 the government forbad anyone to ship rags abroad, under pain of 3,000 *l.* fine. A prohibitive export duty of 30 *l.* per quintal replaced outright prohibition from 1733 to 1771, when the government once again forbad all exports. No one could accumulate stores of rags within four leagues of sea and land frontiers, nor could anyone operate a papermill within the same strip lest the mill serve as a blind for smuggling operations. Only ships of a certain size could accept rags destined for other French ports, and even so the skipper had to post bond that the cargo would reach its official destination. Dutch and English mills offered such high prices for French rags that domestic pro-

ducers always feared that smuggling would deplete the supply available to them. During the first half of the century this concern misled government agencies in certain provinces to restrict the shipment of rags to papermills in other provinces, and the state made the mistake of giving manufacturers exclusive right to buy rags within a radius of one-half league around their mill. Fortunately, an order dated September 10, 1746, rescinded all such monopsonistic rights and removed all restrictions from the shipment of rags from one province to another.

Although the government never conferred as many privileges and tax exemptions on papermakers as it did on entrepreneurs and workers in the glass industry,[5] it exempted papermakers from collecting the *taille*, from billeting soldiers, and from serving in the military. Whereas practically every glass factory of any significance in the eighteenth century received or successfully appropriated the honorific and renumerative title of *manufacture royale*, I have discovered only three papermills which fell in this category. One was an association of several mill owners and paper merchants in Angoumois, which received the title in 1734 but lost it in 1753, after other manufacturers protested. During the last quarter of the century, when Trudaine and the Council of Trade were sparingly using the title to honor exceptional enterprises, the mills at Montargis in Orléanais and at Annonay in Vivarais were singled out. The government also ennobled Pierre Montgolfier at Annonay and awarded gold medals to his two sons because of their contributions to papermaking and also because of their experiments with hot-air balloons. After 1760 and until the Revolution various provincial intendants or parliaments awarded prizes, loans, or monetary gifts with increasing frequency to entrepreneurs who wanted to install Dutch cylinders or to perfect particular processes. The central government also appointed experts to test new devices and seriously attempted to diffuse improved techniques throughout the industry. This was exactly the reverse of the policy it followed during the first half of the century when it tried to freeze standards in production and to prevent experimentation.

Most of the evidence I have sketched in this paper reveals a direct correlation between the industry's expansion and the relaxation of direct governmental controls. After recovering from a long period of depression at the end of the seventeenth century, papermaking perhaps doubled its output between 1720 and 1789, with most of the growth coming after 1750. This was also the period when the relatively large-scale capitalistic enterprise became more common and when more and more manufacturers adopted the Hollander. Government officials no

---

[5] Warren C. Scoville, *Capitalism and French Glassmaking, 1640-1789* (Univ. of California Press, 1950), pp. 83-84, 125 ff.

longer enforced the codes which hamstrung enterprise and initiative; instead, central and provincial agents encouraged those who wished to break with tradition. The state sought to conserve the supply of rags for French producers, to remove some hindrances to the internal trade in rags, to protect domestic mills by taxing imports and removing export duties on paper, and to weaken the power which workers frequently wielded over employers. Such policies definitely worked to promote growth, but others—such as taxing output and restricting the freedom of employers to recruit personnel—worked in the opposite direction.

The fault did not lie entirely with Versailles nor with the limited imagination and capabilities of responsible officials, who followed Colbert and who preceded Gournay, Trudaine, and Turgot. Military and fiscal exigencies often had precedence over problems connected with industry in general and with papermaking in particular. Then, too, established producers rather than stake their future on competition and economic freedom often sought government intervention to correct real or imaginary abuses. Many papermakers had even recommended that the government formulate the ill-advised codes of 1727, 1730, and 1739. The fundamental trouble, if I interpret this segment of French economic history correctly, was identical with that which arises whenever any government attempts to make the basic economic decisions for its people rather than confine itself to formulating policies and general rules within which the economy can function. Its actions are much more likely to be inconsistent and to reflect the immediate concern of those individuals and vested interests for whom economic and technological progress has wrought hardship.

NEW LIGHT ON A STATISTICAL DARK AGE:
U.S. REAL PRODUCT GROWTH BEFORE 1840*

By Paul A. David

Stanford University

"... I would fain advance naught but substantiated facts. But after embattling his facts, an advocate who should wholly suppress a not unreasonable surmise, which might tell eloquently upon his cause—such an advocate, would he not be blameworthy?"—Herman Melville

I

Prevailing scholarly opinion evinces disturbing symptoms of latent schizophrenia on the subject of economic growth in the antebellum United States. Virtually every student of the era now appears quite convinced that the long-term rate of growth of aggregate product per capita underwent a significant acceleration, "somewhere during the period between 1815 and 1860," as Douglass North [8, p. 75] recently phrased it. And, if many economic historians protest Walt Rostow's labeling the years following 1843 as the period of the U.S. "take-off" [11], most do not seem to balk at accepting the guess hazarded by Raymond Goldsmith [4, pp. 277-78]: " a fairly sharp break in the trend of real national product per head . . . occurred not very long before 1839," quite probably during the 1830's. To the uninitiated, it might seem obvious that everybody also subscribes to an explicit set of assertions about the secular rate of growth of per capital product during the period from, say, the founding of the nation to the date of the putative acceleration. Otherwise, how could all agree a significant alteration in the trend had occurred?

Yet, to find any expression of such a consensus on the rate at which per capita product had increased during the half-century following 1790 would be no mean task. Faced with the direct question, most writers draw back from specific statements, pointing to the unfortunate absence of reliable quantitative data describing the secular course of aggregate production during the decades before 1840. The present

lack of satisfactory direct national product estimates, which properly induces such caution, has, somehow, not been allowed to interfere with the general conclusion that the secular pace of per capita product growth changed, becoming notably faster during the latter part of the 1790-1860 interval.

The problem of U.S. growth prior to 1840 has, curiously, come to be treated as a self-contained issue; it is usually split off for consideration in tones much less definite than those assumed in discussions of the (established?) acceleration that took place during the antebellum era. Did output grow faster than the population? How steady was the increase in output? It has been said that these two questions must be answered "if the years from 1800 to 1840 are to be understood as fully as later periods in American economic history" [10, p. 191]. But, can we really grasp the significance of developments observed after 1840 without having to provide those answers for the years before?

This is not to say hypotheses have not been advanced concerning the vexed question of U.S. growth during the 1800-40 interval. Speculation abounds. Indeed, the historiography of the pre-1840 economy, far from a barren field, begins to resemble nothing so much as a Confucian garden—or should one say a garden of confusion—where "a hundred flowers bloom . . . a thousand thoughts contend." These varietals spring up overnight in the space recently cleared by intensive critical examination of Robert F. Martin's [6] estimates of real national income. Martin's figures for the first half of the nineteenth century had real income growing more slowly than the U.S. population in each of the three decades prior to 1830, with per capita real product beginning to rise thereafter, but not reaching as high a level as prevailed in 1800-10 until the 1840's. So strenuous were the doubts subsequently raised concerning the implications of those findings, and the methods by which they were obtained, that the Martin estimates are now generally discarded as a basis for serious discussion of U.S. economic growth before 1840.

Martin's leading critics, however, were unable to come to an agreement on a more believable picture of the course of change during these decades. Without venturing direct estimates of his own, Simon Kuznets [5] argued that a rise in per capita real national income was considerably more plausible than a decline, in view of the increased proportion of the population in gainful occupations and the relative transfer of workers out of agricultural pursuits that occurred between 1800 and 1840. William Parker and Franklee Whartenby [10], on the other hand, questioned the supposition that labor productivity in agriculture increased, or at least remained constant, during this peri-

od; reviewing the available evidence, they judged it "too weak to support Kuznets' inference that per capita real income followed a rising trend from 1800 to 1840."

Douglass North [7] [8] approached the problem from a rather different direction. Shifting attention away from developments impinging immediately upon agricultural productivity, and directing it toward the sharp alterations that took place in the condition of U.S. foreign commerce in the period 1790-1815, North suggested that the years of Neutral Trade (1793-1807) formed an era of "extraordinary prosperity," even in comparison with the period of peace after 1814. Much the same view has recently been propounded by George R. Taylor [12], save that the post-Embargo Act decline is made to appear somewhat less severe, or the subsequent recovery rather more rapid than the movements envisioned by North. Whereas North concludes that "per capita real income was higher in 1799 than it was to be again for a half-century" [7, p. 387], Taylor writes [12, p. 440]: "My own estimate is that the average level of living in 1799-1806 was not reached again until the early 1830's at the earliest, and . . . the average for 1836-40 was at best not much higher than that for the prosperous years around the beginning of the century."

Thus the matter presently stands. Martin's estimates have been banished from sight, but without any specific figures having replaced them, an after-image of a more or less stagnant level of per capital real product still lingers on.

This impression, and the explanations advanced in its favor, are, I suggest, barely more substantial than the Cheshire cat's smile. The evidence that can now be brought to bear upon relevant considerations of a thoroughly familiar kind impels me to advance several counter-propositions about U.S. economic growth prior to the Civil War. For the sake of exposition, these may be stated as robustly as good conscience will allow and rather more categorically than the presently available evidence warrants.

To begin with, the level of real gross domestic product per capita neither declined nor remained stagnant over the 1800-40 period. According to the figures shown in Table 1, per capita real GDP increased during this interval at an average annual rate exceeding 1 percent but probably less than 1.5 percent. While the rate of advance was thus slower than that attained after the Civil War, the gain was hardly insignificant; by 1840 per capital real GDP was probably 60 percent higher than it had been at the beginning of the century.

The growth of real GDP per capita did not, however, proceed steadily during these decades, just as it was not stable over the course of the 125 years that followed. Within the first 70 years of our existence as a

TABLE 1

CONJECTURAL AND MEASURED RATES OF GROWTH OF REAL GDP AND
REAL GDP PER CAPITA IN THE U.S., 1800-60

(Average Annual Percentage Rates)

Intervals*	Real GDP Per Capita (1)	Real GDP (2)
	Long-Term Trend Rates of Growth:	
1800-60‡..........	1.27 (1.38)†	4.32 (4.42)†
1800-35‡..........	1.22 (1.35)	4.28 (4.48)
1835-55..........	1.30	4.40
	Rates of Growth over 20-year Intervals:	
1800-20§..........	0.24 (0.27)	3.29 (3.31)
1810-30§..........	1.54 (1.75)	4.53 (4.74)
1820-40‡..........	1.96 (2.30)	4.96 (5.28)
1830-50‡..........	1.37 (1.51)	4.39 (4.53)
1835-55..........	1.31	4.39
1840-60..........	1.59	4.73

* Initial and terminal dates refer to U.S. census periods; e.g. census year 1799 or calendar year(s) 1799-1800, census year 1859 or calendar 1859-60.
† Parentheses indicate the underlying conjectural estimate(s) is drawn from the Variant I series in Table 2. Unbracketed conjectural rates refer to estimates from Variant II series, also in Table 2, col. 4.
‡ Indicates the initial date observation is a conjectural estimate.
§ Indicates both initial and terminal observations are conjectural.
SOURCE: Col. 1—Table 2, cols. 4 and 5, for chained indexes of real GDP per capita in 1840 and 1860 prices. Col. 2—Underlying real GDP index derived from Table 2 and population statistics in *Historical Statistics of the U.S.* (1960), p. 7, Series A-2.

nation, there were, I suggest, three distinct episodes of accelerated growth, in the course of each of which per capita real product may have increased at rates exceeding 1.5 percent per annum, if not 2 percent per annum.

The first of these movements extended from the early 1790's into the opening decade of the century. In all likelihood, the boom had already begun to slacken before it was brought to an abrupt end by the disruption of the U.S.'s opportunities for continued large-scale participation in international commerce and shipping. As North and Taylor maintain, the heights of prosperity to which the expansion of neutral trade had carried the economy could not be sustained after 1807. But, the ensuing decline hardly could have been as drastic as they have been led to suppose; the sharp deterioration of the international terms of trade, to which North has drawn attention, might have been responsible for something like half of an estimated 10 percent total decline in per capita real income, appropriately defined, between 1800 and 1810. Of course, were we to insist on following conventional real product accounting procedures, the estimated fall would be even less serious, for the 1800-10 deterioration of the terms of trade would have the statistical effect of raising the measured level of real product in 1810 rela-

tive to that in 1800. But, regardless of the way terms of trade effects are treated, it appears that the 1800 level of real GDP per capita had been regained, if not surpassed, before the crisis and depression of 1818-19.

The second surge of accelerated growth, getting under way in the early 1820's, thus came after a protracted period which may well have seen no permanent advance in per capita real income above the level attained when the century was new. This second expansionary movement was no mere recovery from a business cycle downturn; it seems to have lasted for a decade and a half—far outliving the Era of Good Feelings which had attended its initiation—before yielding to marked retardation. Direct estimates of real GDP per capita indicate an average annual growth rate in the neighborhood of 0.6 percent per annum during the decade from the mid-1830's to the mid-1840's, whereas an average growth rate as high as 2.5 percent per annum may have been sustained during the 1820's and early 1830's. (Compare Table 2.)

It is, consequently, quite unwarranted to suppose the accelerated growth of per capita real product observed from the mid-1840's to the late 1850's was unprecedented in U.S. experience or constituted the continuation of a truly secular movement whose origins reached back into the 1830's. This third surge of growth—so much discussed in recent years—had already lost considerable momentum before the outbreak of civil conflict, like the disturbed international conditions of 1807-14, brought in its wake an attenuated period of retarded per capita real product growth.

The instability of the rate of per capita product growth during the antebellum era makes it a treacherous business to infer secular trends from observations drawn at arbitrary census-year intervals. We may attempt to abstract from such variations by considering the trend between 1800, the mid-1830's, and the mid-1850's, occasions on which, it is generally agreed, the U.S. economy experienced strong pressure upon existing capacity and rising prices generated by high levels of effective demand. When this approach is adopted, very serious doubts arise regarding the occurrence of any "sharp break in the trend," let alone an 1843-60 "take-off," or the upward shift "not very long before 1839," which North and others would seem to favor.

Instead, it appears more reasonable to maintain that no significant acceleration of the secular trend in real GDP per capita took place within the era of our national history preceding the Civil War. The evidence now points to average annual rates of increase very close to 1.3 percent in the two trend intervals, 1800-35 and 1835-55, as well as over the whole time span from 1790 to 1860.

II

The foregoing statements, I am well aware, are provocative if not rash. It must be stressed that they do not rest entirely on detailed direct estimates of real product in the U.S. before 1840 comparable with those Robert Gallman [3] has recently published for the years following 1834. Their basis is, nonetheless, a concrete set of "conjectural estimates" (Table 2) arrived at by formally drawing together fragments of quantitative evidence that have accumulated since Kuznets first addressed himself to the problem of U.S. income growth before 1840.

In view of the remaining limitations imposed by the available data, it is still convenient to adhere to a familiar general approach to the problem. Making use of a simple identity, an index of output per capita may be represented as the product of two other indexes: the first describing the relative level of per capita labor inputs and the second the relative level of output per unit of labor input, or, using the labor input measure adopted here, average real product per worker. The problem of estimating the relative levels of per capita product in any pair of years can thus be decomposed into the tasks of arriving at consistently defined estimates of these two underlying magnitudes. Within this general framework it is comparatively easy to marshal evidence bearing directly upon the major empirical issues raised by the controversy over the 1800-40 period. Yet, the provision of answers to questions about the impact of the changing industrial distribution of the labor force upon average labor productivity in the economy, or the effect of the westward movement upon labor productivity in agriculture, is treated here as subsidiary to the task of deriving plausible estimates of per capita real product. And that problem can best be regarded as one of controlled prediction or "retro-diction" rather than as an exercise in the analysis of observed phenomena.

The particular prediction model within which available scraps of information relating to the period 1800-40 can be readily assembled is, in its general features, extremely simple, almost too naïve to claim so fancy a title. Specifically, (a) the economy is sectored into agricultural pursuits and all other productive activities; (b) the intersectoral shift component of overall labor productivity change is gauged with reference to the relative levels of output per worker in the farm and nonfarm sectors at one point in the period, the 1839-40 census year; (c) it is assumed that the (weighted) average intrasectoral rate of labor productivity growth in both sectors equaled the rate of growth of output per worker in the dominant activity, agriculture. With these speci-

TABLE 2

TRENDS IN REAL GROSS DOMESTIC PRODUCT PER CAPITA, U.S., 1800–60

| Year* | Basic Data Used in Calculation of Col. (4): Indexes, 1840=100 | | | Real GDP Per Capita Indexes, 1840=100, Based on: | | |
| | | | | Conjectural Estimates in 1840 Prices (4) | | Direct Estimates in 1860 Prices (5) |
	sA_t (1)	$\dfrac{(\pi_A)_t}{(\pi_A)_o}$ (2)	$\dfrac{(\rho)_t}{(\rho)_o}$ (3)	Variant II†	Variant I‡	
1800........	.826	76.4	96.6	64.4	60.2	
1810........	.837	73.9	97.0	61.9	57.8	
1820........	.790	76.6	98.7	67.6	63.6	
1830........	.707	90.0	97.6	84.0	81.8	
1835........		...	...	98.6	98.6	98.6
1840........	.634	100.0	100.0	100.0	100.0	100.0
1845........		...	...	"Test Predictions":		104.5
1850........	.548	100.0	106.5	113.0		110.4
1855........		...	...			128.0
1860........	.532	121.5	106.5	138.0		137.0

* Census year intervals; e.g., 1799–1800, 1809–10, etc.
† Based on Definition II measure of δ= .511.
‡ Based on Definition I measure of δ= .400.

SOURCES: The sources given here do no more than identify the data by references to the tables (each accompanied by extensive notes and sources) in my forthcoming article, "U.S. Real Product Growth Before 1840: New Evidence, Controlled Conjectures," in the June, 1967, issue of the *J. of Econ. Hist.*

Columns 1, 2, 3: The variables are those defined in the text above. For Col. (1) here, cf. Table 4, col. 1; for Col. (2) here, cf. Table 6, line c. 2 for 1800–40, and Table 2, col. 3, for 1840–60; for Col. (3) here, cf. Table 3, col. 1.

Column 4: Computed, following the formula given in the text above, from Cols. (1), (2), (3) here, and the alternative estimates of the relative level of agricultural output per worker, given in notes (†) and (‡) to this table.

Column 5: Cf. Table 8, col. 3.

fications, and the available direct estimates of the overall rate of U.S. labor force participation, it is extremely easy to calculate hypothetical or, as I prefer to call them, conjectural indexes of real gross domestic product per capita.

Estimates of per capita real output (v), identical to those implied by the foregoing general specification, can be computed on the following more restrictive assumptions: (1) the ratio of agricultural output per worker π_A, to nonfarm output per worker, π_N, remains constant through time; (2) since $(\pi_A/\pi_N) = \delta$, is a fixed parameter, the average intrasectoral labor productivity trend is identical to the trend of output per worker in either of the sectors; i.e., $(\pi_A)_t/(\pi_A)_o = (\pi_N)_t/(\pi_N)_o$, by assumption.

Define real GDP per capita as: $(V/P) = v$; the farm labor force as a proportion of the total labor force as: s_A; the ratio of the total labor force to the population as: ρ. For any given date, we may then write the identity,

$$v = [s_A \pi_A + (1 - s_A)\pi_N]\rho.$$

And, since it is assumed that $(\pi_A/\pi_N)_t = \delta$, for all dates, the desired index may be computed from:

$$\frac{v_t}{v_o} = \frac{(\rho_t)(\pi_A)_t[(1/\delta) - (1/\delta - 1)s_A]_t}{(\rho_o)(\pi_A)_o[(1/\delta) - (1/\delta - 1)s_A]_o},$$

given observations of the variables ρ, π_A, and s_A.

Simple assumptions fail to do justice to the complexities of long-run economic change, and naïve, highly aggregated models may not carry us far towards a satisfactory understanding of the way the behavior of producers and consumers meshed together to yield the historical course of events observed in a market economy. Yet, when it comes to predicting or simulating aggregate phenomena, it is quite frequently found that an admixture of naïveté and aggregation is not necessarily bad.

Consider the performance of the present prediction equation in describing the growth of per capita real GDP in the U.S. over the course of the two decades following 1840. The naïve calculations suggest that a 38 percentage point increase occurred between 1840 and 1860. (Cf. Table 2, col. 4, Variant II.) Independent direct estimates of GDP per capita in constant (1860) prices—derived from Gallman's recent work (cf. Table 2, col. 5)—show a 37 percentage point rise between those two dates. For the shorter, intercensal periods, the predictions are only slightly less impressive: they overstate the percentage change "observed" during the 1840's by 3 percentage points and understate the "observed" change in the 1850's by 2 percentage points. These, re-

member, are predictions of proportional changes in per capita real GDP; the conjectural estimates of the levels of real GDP fall still closer to their respective marks. I daresay Professor Gallman himself would not think it reasonable to strive for greater accuracy, in view of the margins of error he attaches to his real GNP estimates [3, pp. 52-57, 62].

There is, of course, no reason to suppose the present prediction model would prove universally dependable; in an important sense, its application in this particular context does represent a judgment that the formulation is sufficiently flexible to capture two features of U.S. antebellum development which exercised a decisive influence upon the long-term growth of aggregate labor productivity and per capita real product. I take these features to be the continuing predominance of agricultural employment and the radically expanding relative importance of nonfarm employment and production. For the moment, the test results obtained for the 1840-60 interval on the basis of that judgment are sufficiently heartening to tempt inquiry into what the available data imply concerning the years before 1840.

III

On this occasion, an adequate presentation of the evidence I have assembled, including a detailed description of the new labor force and labor productivity measures which have been prepared in connection with this study, would scarcely be feasible. Adherence to the requirement of brevity permits no more than a series of bald statements of findings on the subsidiary questions around which the controversy over the pre-1840 period has swirled.

The first four points of interest relate to the growth of the labor force and its changing sectoral distribution during the 1800-40 interval.

1. Although between 1800 and 1860 the overall U.S. labor force participation rate rose 10 percent, from .32 to .35 approximately, the increase that took place over the decades prior to 1840 amounted to only a 3.5 percent gain. (Table 2, col. 3.)

2. If the course of change in labor force participation offers scant support for the view that per capita real product rose substantially during the 1800-40 interval, just the opposite must be said of the impressive structural change brought to light by the same set of labor force estimates. The proportion of the labor force engaged in nonfarm pursuits expanded from .17 in 1800 to .37 in 1840. (Table 2, col. 1.)

3. At the time of the Sixth Census (1840), average gross farm product per worker was four-tenths to one-half as large as average gross domestic product per worker in all nonfarm activities. To avoid over-

stating the intersectoral shift effect on average labor productivity, major reliance is placed on the higher ratio (0.511); it indicates a relative level of farm labor productivity more than twice as high as the figure Kuznets employed in his intersectoral shift calculations for the U.S. in the period between 1800 and 1840 [5, pp. 222-24].

4. Average labor productivity in the nation, it is estimated, increased by 16 (if not by 22) percent as a consequence of structural changes reflected in the altered sectoral distribution of the work force over the period 1800-40. This, combined with the small rise in labor force participation, would have raised real GDP per capita by at least 18.6 percent, making no allowance for the growth of average intrasectoral productivity.

What, then, can be said regarding the course of agricultural labor productivity before 1840? That is, after all, the key question. For, *pace* Professor North, it is readily shown that the trend in labor productivity within the agricultural sector must have exerted a dominant influence upon average intrasectoral productivity growth for the entire economy and, hence, upon the trend in per capita real product. Examination of the evidence now at our disposal leads me to hypothesize—for the purpose of the present calculations—that although the first two decades of the nineteenth century may well have passed without any notable advance having occurred in U.S. agricultural labor productivity, by 1840 real farm product per worker was substantially higher, perhaps 31 percent higher, than it had been c. 1800. Five supporting points should be quickly noted under this heading.

1. The foregoing pattern of farm labor productivity gain is implied by estimates of farm gross product, in constant 1839-40 prices, which are largely based on the assumption that the volume of farm production increased no faster than the nation's population. (Table 2, col. 2.) For per capita consumption of all farm products to have remained strictly constant, real farm output per worker would have had to have increased by at least 26 percent between 1800 and 1840.

2. Note that this finding implies that a reduction in agricultural labor productivity sufficient to offset the incipient rise in per capita real product over the 1800-40 period on other counts—i.e., increasing labor force participation, structural change, and labor productivity growth in the nonfarm sector—would have entailed a reduction in per capita agricultural output by more than 48 percent. One must be prepared to swallow improbabilities of that dimension in order to passively subscribe to the idea that the level of U.S. real GDP per capita was much the same in 1840 as it had been when the century opened.

3. A 31 percent overall rise in agricultural output per worker is certainly closely in line, with the labor productivity trend estimates for

major farm crops, such as wheat, cotton, and corn, during the 1800-40 interval. The estimates to which I refer are those made by Cooper, Barton, and Brodell [1, p. 3] of the U.S. Department of Agriculture. Still, since doubts have been expressed [10, p. 207] as to the accuracy of these USDA trend estimates, we ought not rely wholly upon the corroboration they provide for the view that agricultural productivity was rising.

4. If one is willing to infer general trends in the marginal physical productivity of farm labor from data on real labor costs, a fine resemblance is to be observed between the broad movements in marginal productivity inferred from the latter, and the course of change suggested by our conjectural estimates of average labor productivity in the farm sector. My estimates of the marginal productivity of free farm laborers engaged in northern agriculture indicate something between a 26 percent and a 50 percent increase over the period 1800-40, the entire advance having occurred after 1820.

5. Finally, quantitative evidence now available from the work of Richard Easterlin [2] and William Parker and Judith Klein [9] discloses that at the time of the 1840 Census, physical labor productivity was considerably higher in the new, interior farming areas than it was on farms and plantations located in the eastern seaboard states. Thus, a calculation of the effect of the regional redistribution of the farm labor force during the period 1800-40 indicates the westward movement had most probably raised U.S. average farm labor productivity rather than lowering it as some writers [10] [12] have suggested. The estimated gain on this account, although biased toward understating the true magnitude of the "regional shift effect," is approximately 8 percent, or roughly one-fourth of the conjectured improvement in agricultural labor productivity during the forty years in question.

As a basis for making conjectural estimates of per capita real GDP, it has been assumed that the average intrasectoral labor productivity increase—for the farm and nonfarm sectors—followed the course of labor productivity within the farm sector. The indirect implications flowing from that assumption suggest it may not be too far from the truth, and that if it is incorrect, the direction of the resulting errors will be towards understating the long-term rise of per capita real product.

1. First, the available evidence relating to the real costs of farm and nonfarm labor—imperfect as it is—suggests that the rise in farm labor productivity prior to 1840 proceeded more slowly than the increase in output per worker in manufacturing, transportation, and other nonagricultural activities.

2. The implied total increase in average U.S. labor productivity over the period 1800-40 works out to be about 52 percent. Slightly less than a third of the total change is "attributable" to the structural change reflected in the altered distribution of the labor force. If this is purely a conjecture, at least it is not an outlandish conjecture; the relative importance of the intersectoral shifts' "contribution" to labor productivity growth is quite similar to that observed in the U.S. during the latter half of the nineteenth century.

3. Finally, the conjectural estimates of per capita real GDP to which we are led (in Table 2, col. 4) simply say that if farm production essentially kept pace with U.S. population growth between 1800 and 1840, real gross domestic product in the nonfarm sector increased 9.1-fold, or only 2.8 times as fast as the nation's population. Considering the spectacular rates of expansion of industrial production recorded in this period—such as the 10-fold rise in the volume of New England textile output achieved in the brief interval from 1826 to 1840 —the 55 to 62 percent rise in per capita real gross domestic product implied by these estimates would appear a reasonably modest appraisal of the economic progress achieved in the United States during the first four decades of the century.

IV

For those who would maintain that the onset of modern economic growth in the U.S. was signaled by a decisive break in the long-term trend of per capita real product, the evidence now must pose something of a paradox, if not a bitter dilemma. No such secular acceleration can be found within the antebellum era. The present estimates, instead, place the rate of growth of per capita real GDP near 1.3 percent per annum for both the periods 1800-35 and 1835-55, as well as for the entire interval from 1790 to 1860. On the other hand, after the Civil War the long-term growth rates of per capita product were higher than this, being closer to 1.8 percent per annum than to 1.3 percent per annum. Furthermore, Gallman's [3, p. 11] most inclusive estimates of the real gross domestic capital formation ratio in the U.S. also suggest a more notable upward drift between the ante- and postbellum periods than had occurred within the years between 1834 and 1860. And yet, it was during the thirty years before 1850, not in the thirty years after, that an impressive transfer of labor into nonfarm pursuits, among them factory employments, was accomplished.

Are we then to conclude there was no true U.S. "take-off" in the antebellum era, and plunge immediately into the diverse histories of American colonies in the hope of discovering some prior fundamental "break in the trend"? Or, attempt to refurbish with cast-off Rostovian

garb the now very unfashionable idea of the industrial revolution "taking hold" in the U.S. only after the Civil War? At present, it seems a far more sensible solution would be to abandon the notion that significant, portentous stirrings of urban-industrial development within the predominantly agrarian antebellum economy would quickly be reflected in a discontinuous and permanent alteration of the per capita real product growth rate or in a parallel upward shift of the aggregate capital formation ratio.

British economic historians now freely entertain the hypothesis that some time elapsed before the structural transformations associated with rapid industrialization made their impact felt throughout the entire British economy; and that while a discontinuity can be seen in the long-term rate of industrial output growth late in the eighteenth century, the ensuing acceleration of per capita real income growth was a more gradual affair. Is it not fit that in this, as in other things, we might fruitfully follow a British lead?

REFERENCES

1. Martin R. Cooper, Glen R. Barton, and Albert P. Brodell, *Progress of Farm Mechanization,* U.S. Dept. of Agriculture Misc. Pub. 630 (Washington, Oct., 1947).
2. Richard A. Easterlin, "Estimates of Regional Agricultural Output Per Worker for the U.S. in 1840, in Standard Agricultural Output Prices" (unpublished estimates, kindly made available by Professor Easterlin in private communication).
3. Robert E. Gallman, "Gross National Product in the United States, 1834–1909," in *Output, Employment, and Productivity in the United States After 1800,* N.B.E.R. Studies in Income and Wealth, Vol. 30 (New York, 1966), pp. 3–76.
4. Raymond W. Goldsmith, Hearings Before the Joint Economic Committee, U.S. 86th Congress, First Session, Part 2—*Historical and Comparative Rates of Production, Productivity and Prices* (Apr. 7, 1959).
5. Simon S. Kuznets, "Long-Term Changes in the National Income of the United States of America Since 1870; Appendix—Current National Income Estimates for the Period Prior to 1870," *Income and Wealth of the United States, Trends and Structures,* ed., Simon S. Kuznets (Cambridge, 1952), pp. 221–41.
6. Robert F. Martin, *National Income in the United States, 1799–1938,* Nat. Ind. Conf. Bd. Studies, Number 241 (New York, 1939).
7. Douglass C. North, "Early National Income Estimates of the United States," *Economic Development and Cultural Change,* Apr., 1961, pp. 387–96.
8. ———, *Growth and Welfare in the American Past, A New Economic History* (Englewood Cliffs, N.J., 1966).
9. William N. Parker and Judith L. V. Klein, "Productivity Growth in Grain Production in the United States, 1840-60 and 1900-10" in *Output, Employment, and Productivity in the United States After 1800,* N.B.E.R. Studies in Income and Wealth, Vol. 30 (New York, 1966), pp. 523–80.
10. William N. Parker and Franklee Whartenby, "The Growth of Output Before 1840" in *Trends in the American Economy in the Nineteenth Century,* N.B.E.R. Studies in Income and Wealth, Vol. 24 (Princeton 1960), pp. 191–212.
11. W. W. Rostow, *The Stages of Economic Growth* (Cambridge, 1960).
12. George R. Taylor, "American Economic Growth Before 1840: An Exploratory Essay," *J. of Econ. Hist.,* Dec., 1964, pp. 427–44.

DISCUSSION

STANLEY ENGERMAN: Paul David's paper represents the most thorough study in the debate, set off by the income estimates of Martin and their detailed criticism by Kuznets, concerning American economic growth between 1800 and 1840. It is based both on new time series not previously utilized and upon the resourceful application of available data to clarify certain key points. The new time series are Lebergott's reworking of labor force data, total and by industrial distribution (usefully revised for certain years by David), and the Towne-Rasmussen series on nineteenth-century agricultural output. David has also prepared 1790 estimates of the labor force and agricultural output linked to these two series. The points clarified concern the effects of the westward movement and foreign trade upon the course of per capita income. David uses Parker's crop estimates and a heretofore unpublished set of estimates prepared by Easterlin showing 1840 agricultural output by regions in uniform national prices. With them, he answers Parker-Whartenby and Taylor concerning possible adverse effects of the westward movement. Similarly, his use of North's data on foreign trade and the terms of trade permits him to strongly argue that North and Taylor have considerably overstated the case concerning the effect of trade on income movements.

David's conjectures are used to argue against two hypotheses concerning economic development in the first half-century of nationhood. Because he is attacking conflicting hypotheses, David's discussions of possible biases in the results are somewhat misleading. What is a bias against one conclusion can be a bias in favor of the other. However, a discussant is placed in the comfortable position of just raising questions without concern about their effects upon the outcome of the debate. The first attack is on the conclusion, accepted by Martin, North, and Taylor, that per capita income was roughly the same in 1840 as it was in 1800. David marshals strong evidence to argue that such an occurrence was unlikely. The reasons are those first suggested by Kuznets—a shift within the labor force to the higher productivity (per worker) nonagricultural sector and a rise in the ratio of labor force to population. The relative importance of these differ in David's estimates, based as they are upon a different set of labor force estimates. Thus even if labor productivity within each sector had remained constant, per capita income would have risen; and David uses evidence based on wage and price data to support the claim of intrasectoral labor productivity increases.

The second argument is with those such as Rostow and Goldsmith who assert that there occurred a sharp acceleration in the rate of growth of per capita income at some time in the antebellum period. David claims that there was no such acceleration between 1790 and 1860, and suggests that no upsurge occurred prior to that period. On this point, although my intuition would be to believe no sharp break did occur, less confidence can be placed in inference based upon the income estimates. If we eliminate the decade of the 1790's, acceleration in the growth rate did occur after 1820, with the long swi

pansion of 1820-34 being one of the most rapid periods of growth in U.S. history. (It should be noted that this is the period stressed by Fogel in his critique of Rostow.) However, David points to a high estimated growth rate between 1790 and 1800, making the 1800-20 interval a lull between two periods of rapid growth. Given the heavy reliance upon the structural change in the distribution of the labor force for the 1790's estimate, one would prefer stronger supporting evidence. Such evidence is undoubtedly unavailable, but the magnitude of the shift into nonagricultural employment seems high in light of the not abnormal urbanization of that decade. David's estimates imply a growth rate of nonagricultural output of about 10 percent per annum, which must be among the highest rates recorded and exceeds that in any pre-1860 decade. And given the agricultural labor force shares shown in the time series and cross-sections collected by Kuznets, if the 1790 estimates are correct, it is difficult to visualize any such marked growth rates in the pre-1790 period.

There are several surprising implications in the conjectural estimates which, if true, will force revisions in the way we look at the pattern of U.S. development. Since the Towne-Rasmussen estimates for 1800-40 rest almost entirely upon the assumption of constant per capita domestic consumption, the export figures having only minor effect, there is no trend in per capita agricultural output before 1840. Gallman's data for 1840-1900 show only a small increase in per capita agricultural output, the one irreversible rise occurring in the decade of the 1870's. Thus, abstracting from fluctuations, the data suggest about a zero income elasticity of agricultural output per capita for a period of 110 years. The Lebergott series, which in a sense is the key to David's findings, presents a much slower growth in the agricultural labor force from 1800 to 1840 than do the series used by previous writers; 2.4 percent per annum versus, e.g., 3 percent in the series used by Towne-Rasmussen. It is the relative decline in the share of the labor force in agriculture with the assumption of constant per capita consumption which yields David's estimates of increased output per worker in agriculture and nonagriculture. (Use of earlier agricultural labor force estimates in conjunction with Lebergott's participation rates would show almost no measured growth in per capita output. Given our proclivity to justify what data appear to reveal, perhaps some tolerance should be granted those who used the earlier figures.) While the Lebergott series is clearly superior to earlier variants, it does imply that a substantial part of the nineteenth-century shift out of agriculture was concentrated in the period prior to 1850, the years from 1850 to 1880 showing a small decline. In David's estimates there results an implied rate of growth of agricultural output per worker from 1810 to 1840 in excess of the rate of increase for the rest of the century! The 1820-30 increase in agricultural productivity (a decade in which ___'s data show a small contribution of the westward movement) ___he peak decades of the later part of the century. Given ___cultural mechanization, these are surely surprising infer-

___ interpreting David's description of the procedure used to ___ estimates. David computes the 1840 relative sectoral pro-___er and, for computational purposes, holds this ratio con-___o reduce the measured effect of intersectoral shifts he uses

the variant with the higher agricultural relative. However, he then states that his other decade estimates are based upon the less inclusive concept of GDP. While one can understand his desire to provide a unique measure of the effect of the intersectoral shift, as well as his attempt to minimize that effect, his procedure is misleading. Based on his formula, all that matters are changes in the variables, the 1840 absolute level being irrelevant except to fix relative sectoral productivities. Once this ratio is fixed, consistent totals of sectoral outputs can be derived only by using that variant used to set the ratio. The use of another variant builds in an implicit trend in relative sectoral productivities which is not obvious, and here implies a relative increase in product per worker in agriculture after 1800. The two concepts discussed by David did grow at different rates, but that does not justify the procedure, and the concept David used should yield higher growth rates than would the more inclusive measure. While estimates permitting changes in relative output per worker in the different sectors may be better than those assuming a constant ratio, it would seem better to have such assumptions explicitly introduced.

Finally, how comfortable can the historian be with David's conjectural estimates? While the extrapolation of certain ratios may permit the setting of bounds and subsidiary data indicate the plausibility of such bounds, the procedure biases the conclusion for those who would argue that there were discontinuities in the process of growth. David's results rest heavily on the assumptions of near constant per capita agricultural output between 1790 and 1840 and a constant ratio of output per worker in agriculture relative to non-agriculture (at the 1840 level). For those who argue that these ratios fluctuated or changed, no matter which conclusion they wish to defend, David has put aside much of the controversy. When David notes the success of his prediction formula for the 1840 to 1860 period, he is basically showing the stability of one of the ratios, relative sectoral productivities, but since he is using actual output estimated by Gallman, the Towne-Rasmussen constant per capita consumption assumption is not tested. The latter would seem as crucial an assumption as the one tested by David.

All of these comments do not do justice to the meticulous care with which David has prepared his estimates, justified them, and analyzed the literature. His conjectures will form a necessary starting point for the further analysis of these early years of economic development in the United States.

There are two points I should like to raise about Professor Scoville's interesting paper that are relevant to any study of one particular industrial sector. First, what was happening to the size of the market for the industry? Perhaps there is no data available to determine the overall growth of the French economy, but it would be interesting to determine the effects of the loss of foreign markets due to tariffs. If Coleman's contention about the sheltered nature of the English industry is correct, it may be that the relative difficulties of the French industry are overstated. Second, Scoville's discussion of the disadvantages of government interference and the presumed beneficial effects of liberalization may be open to some question. Such liberalization as did occur was still within the confines of a protective tariff and, perhaps more importantly, heavy and apparently increasing control over the raw material, rags, which formed a large proportion of total costs. The export of rags was prevented, ei-

ther by explicit provision or prohibitive duty, and the control over supply led to restrictions upon the location of raw material stores and even the location of the paper establishments themselves. Perhaps the implication is that appropriate government policies of control over inputs can lead to rapid growth of final output of an industry, since it would seem that people dealing in rags did not benefit from any liberalization.

Ross M. ROBERTSON: These papers have been particularly rewarding, largely because for the first time in several years we have listened to essays in economic history rather than about economic history. These papers suggest that there are not really two kinds of economic history, as frequently alleged, but rather just one, based on solid scholarship and aimed at finding out the truth about the past.

As Professor Scoville suggests, anyone familiar with Professor Hamilton's work knows that he is not afraid of marshaling data. (The British, incidentally, are still arguing about conclusions he reached in his seminal work published more than thirty years ago.) Scoville, though cautioning us about the validity of his own data, nevertheless gives us a good notion of quantitative change in the French paper industry during the eighteenth century.

Yet each of these first two papers deals with ideas rather than quantities. Hamilton devotes a large portion of his paper to two points. The first, that John Law was not a compulsive or professional gambler, is a result that we might have expected. The second, that Law was a vindictive man, continually exhibiting hostilities that would bring him to ultimate ruin, is a little harder to believe. Yet assuming that Hamilton has chosen adverbs with his usual care, we can only conclude that Law's disintegration began inside the man rather than outside.

I find Scoville's paper on the whole persuasive. I certainly have no quarrel with his quantitative inferences, nor for that matter with his description of French governmental regulation in the eighteenth century. I was not altogether convinced of the direct correlation between the paper industry's expansion and the relaxation of direct governmental controls that he alleges. Perhaps Scoville's more extended publications on this subject will elucidate this relationship.

I must confess to a special interest in Professor David's paper on real product growth in this country before 1840. Like many another middle-aged historian in this room, I recall how puzzled we were when Robert F. Martin first published his income and product estimates in the late 1930's. Even if we allowed for errors and for the fact that some of the census years were depressed and others prosperous, it appeared that per capita income of Americans declined steadily between 1799 and 1829. Converting Martin's series of private production income per capita to an index (1799 = 100), we have:

	1799	1809	1819	1829	1839
Current dollar series.....	100	99	71	58	74
Cost-of-living series......	100	96	79	77	92

From these data it could only be concluded that growth of output was particularly slow during the 1810's and 1820's.

Professor David reports the concern that Kuznets and, later, Parker and Whartenby expressed about the Martin estimates. In formulating his theory of a "take-off" of the American economy during the 1840's, Professor W. W. Rostow based his propositions more on twentieth-century experience than on nineteenth-century evidence. Actually, it was Professor Robert E. Gallman's calculations, published in the same volume with the Parker-Whartenby results, that cast the final shadow on Martin's work, for Gallman's estimates indicated astonishingly high decennial increases in the value of commodity output for the period 1839-59. Indeed, a rate of growth for this period of more than 50 percent in the value of commodity output made it almost impossible to believe that the orthodox view of growth rates between 1789 and 1839 was correct.

Basing my calculations on changes in the balance sheets of selected financial institutions, I wrote some years ago that income per capita may have increased by as much as two-thirds over the fifty-year period 1789-1839.[1] It was possible then to hypothesize that the rate of growth of the American economy did not jump suddenly in the 1840's, but instead accelerated only slightly between 1790 and 1860. Such a hypothesis, if substantiated, meant that there would have to be a revision of our ideas about the way growth took place in the United States.

It seems to me that Professor David's evidence is convincing. He concludes that by 1840 real gross domestic product, per capita, was probably 60 percent higher than it had been at the beginning of the century. With the tendency of the American economy toward pronounced economic fluctuation in the nineteenth century, we can readily accept the observation that the rate of change was not steady. On the other hand, the older historians exaggerated the swings in output experienced in the first half of the nineteenth century. When we reflect that interchangeable-parts and continuous-process manufacture were pre-1800 innovations in this country, that the factory system was introduced in the textile industries before 1815, and that there was tremendous investment in canal building in the 1810's and 1820's, we should not after all be astonished at a more rapid rate of growth than previously suspected. When we reflect further that the commercial banking system was ideally devised to promote growth and that central-bank experimentation, modified by the vulgar profit motive, did little to inhibit free-wheeling early entrepreneurs, we are even more readily persuaded by David's pathbreaking paper.

These findings will of course be refined and modified. I am disposed to think, however, that further inquiry into colonial rates of growth may lead to the conclusion that the American economy grew at approximately the same rate for three hundred years. If it did, some fashionable growth models, even now of dubious value, will have to be changed.

[1] See Ross M. Robertson, *History of the American Economy* (Harcourt, Brace and World, 1964), p. 239.

ECONOMIC DEVELOPMENT

SOME LESSONS OF HISTORY FOR DEVELOPING NATIONS*

By Rondo Cameron

University of Wisconsin and
Instituto de Economia, Universidad de Chile

In recent years economists have recognized that historical experi-
ence can be a valuable guide for both theory and policy [2]. Nev-
ertheless, many economists and policy-makers, especially in developing
nations, persist in believing that the historical experience of modern
industrial nations is irrelevant to the problems faced by developing na-
tions [23] [30]. The purpose of this paper is to argue the contrary
proposition. This is not to say that contemporary developing nations will
inevitably follow the same pattern of development as the highly indus-
trialized nations, much less that the policies of the nineteenth century
are appropriate in the vastly different circumstances of the second half
of the twentieth century. Historians would be the last to deny that there
are alternative routes to economic development, or that different condi-
tions require different policies. Even Marxist economists no longer be-
lieve Marx's dictum that underdeveloped nations can see in the history
of the developed ones the mirror image of their own future. This does
not signify, however, that nothing at all can be learned from history.
There are, in fact, "lessons of history" as abstract from time and place
and almost as ineluctable in their import as a multiplication table.
For policy-makers in developing nations to ignore them—as they do as
often as not—is to condemn themselves to failure and their nations to
perpetual poverty.

The following brief discussion is limited to four areas of major im-
portance for contemporary policy-makers. Limitations of space force
me to omit other areas scarcely less important, such as the role of agri-
culture in industrialization, and to condense drastically the available
evidence on the topics mentioned.

The Role of Capital

Economists from Marx and Engels to Harrod and Domar, as well as earlier and more recent ones, have stressed the preeminent importance of the role of capital in economic growth. After World War II the almost unanimous and sometimes almost the only prescription of economic advisers from developed nations for those beginning to industrialize was massive injections of capital [20] [23] [24]. Rostow's celebrated take-off hypothesis, which emphasized a rapid rise in the rate of net investment from 5 percent or less of national income to 10 percent or more in order to achieve sustained growth, appeared to give empirical support to the theoretical propositions. In a number of countries the prescription of the capital intensive economists was carried out through a combination of capital imports and forced savings by means of government finance. Although a few countries have achieved relatively high rates of growth, the results as a whole have been far from satisfactory; nor is there any clear correlation between investment ratios and rates of growth [29].

Criticism of the take-off hypothesis based on recent historical research has virtually eliminated the key role assigned to the rapid rise in the investment ratio. This research indicates that almost every developed country of today entered a phase of sustained growth with investment ratios substantially below the magic figure of 10 percent; and that the rise in that ratio followed rather than preceded the adoption of new technologies. For Great Britain the work of Phyllis Deane and her collaborators shows that for most of the eighteenth century the rate of net capital formation measured as a percentage of national income did not rise above 5 percent, reaching possibly 7 or 8 percent only in the last decade of the century; not until the railway boom of the 1840's—that is, well after the industrial revolution proper—did the ratio rise to 10 percent [8] [9] [10]. Moreover, according to Miss Deane, during the full flush of the industrial revolution at the end of the eighteenth century "the annual flow of new capital into the leading commercial and industrial sectors . . . was not more than . . . perhaps 1 percent of national income" [8, p. 366]. Agriculture, transportation, and urban construction each absorbed more capital than mining and manufacturing together.

Similar situations appear to have obtained in other countries during the early stages of industrialization, although for most of them the statistical evidence is sparse and unreliable. Preliminary results obtained by Jean Marczewski's team suggest that in France net capital formation as a percentage of net domestic product did not rise above 3 percent per annum, on the average, until the 1840's, and did not rise

above 10 percent until the railway and urban construction boom associated with the Second Empire [19, p. 121]. The 3 percent figure appears to be implausibly low, and is probably due to imperfections in the basic data; but Marczewski himself admits that his estimates for the second half of the century are probably too high.

The earliest figures for Germany pertain to the 1850's, when German industrialization was already in full swing, and indicate a ratio of approximately 7.5 percent [15, pp. 114, 115]. Professor Hoffmann, the leading authority, regards 5 percent as a "realistic figure" for the period 1831-55. In Sweden, the country in which the process of industrialization is best documented statistically, the ratio apparently did not exceed 6 or 7 percent until the twentieth century [5] [16]. For Japan, finally, the best estimates seem to indicate that the ratio of net investment to national income remained roughly constant at 7-8 percent from the 1880's until the early years of the twentieth century; not until after World War I did the ratio rise appreciably [26].

What is the significance of these data for theories of economic growth which assign such great importance to the factor capital? In the first place, it seems clear that the increase in the investment ratio is a consequence rather than a cause of economic growth; rich nations, like rich individuals, tend to save and invest a larger proportion of their incomes. Second, capital is not, as we used to be taught, a homogeneous factor; this seems obvious to economic historians, but it is only recently, as a result of developments on other fronts, that we have become aware of the importance of the heterogeneity of capital. In particular, capital as conventionally measured does not take adequate account of investment in human skill and intelligence. Taking this into account, it seems likely that today's industrial nations began modern economic growth with substantially larger stocks of capital than would appear from the statistical evidence. This is a point to which I shall return. Important as these considerations are, however, they do not explain how economic growth began, nor the precise role of capital in initiating growth. From historical evidence it appears that the key development in getting the growth process under way was a more intensive and efficient use of both the existing stock of capital and of normal increments to that stock.

It is easy to demonstrate theoretically that an economy can experience an increase in total product per worker even with zero net investment by means of the replacement of worn-out capital equipment with new equipment having a higher productivity as a result of technological improvements. Similar results can be obtained as a result of organizational improvements or financial and administrative innovations. Both processes appear to have been at work during the English in-

dustrial revolution and similar periods of rapid change elsewhere. So simple an invention as the spinning jenny—the first of a series of innovations that revolutionized the cotton industry—permitted one man to do the work of sixteen. The adoption of the factory system increased productivity through a more intensive use of capital equipment and by permitting more continuous and efficient supervision of labor. In connection with improvements in transport (turnpikes, canals, railways), it also permitted a reduction in the proportion of working capital required under the domestic system of production. The adoption of steam power and artificial illumination freed early factory industries from reliance on the raw forces of nature and allowed more continuous operation regardless of the season.

Financial innovations associated with the rise of banking, including the use of banknotes and demand deposits as stores of value and means of payment, economized on working capital and permitted a more efficient allocation of savings to investment [7]. Traditional economies typically "lock up" a large proportion of their meager stocks of capital in such unproductive forms as treasure hoards (including hoards of commodity money) and excessive inventories of foodstuffs and other commodities. The introduction of financial intermediaries allowed this "primitive accumulation" to be used more productively by providing holders of idle capital with a wider variety and more attractive range of financial assets, as well as by giving entrepreneurs more convenient access to loanable funds.

The same considerations of intensity and efficiency apply to the use of foreign as to domestic capital. By and large, foreign capital in underdeveloped nations which is not accompanied by the technical and organizational skills to put it to effective use, as is the case with the large majority of loans to governments, is unlikely to make much of a contribution to economic development. The results of the relatively small investment of French capital in Belgium and Germany in the mid-nineteenth century stand in marked contrast to the much larger investments in Spain, Italy, and Russia [6].

What are the lessons in this for contemporary developing nations? Under the influence of outmoded economic theories, policy-makers in many developing nations have developed a naïve faith that a rise in the investment ratio is a necessary if not a sufficient condition for economic development. It should be made clear to them, therefore, that such a rise is neither sufficient nor necessary. While in most cases an increase in the rate of real capital formation would be desirable, it does not follow that development cannot take place by means of a more efficient allocation and use of existing stocks of capital and normal increments thereto. Almost every underdeveloped country has its

examples of idle, underutilized, or misallocated capital: extravagant public buildings and similar monuments, such as athletic stadia; unprofitable "prestige" airlines; large modern factories which remain idle or work at less than full capacity because of poor planning, bureaucratic interference, lack of markets, shortages of raw materials, etc. Most developing nations hold excessive inventories (relative to the flow of total product) of both producers' and consumers' goods as a result of slow and uncertain deliveries from both domestic and foreign suppliers due to poor transport facilities, exchange controls, and administrative delays. Even so, the periodic temporary exhaustion of raw materials and spare parts results in the enforced idleness of much capital equipment. In short, there are still many economies to be achieved. The emphasis in development planning should shift, I suggest, from attempts to raise the investment ratio at whatever cost, to attempts to ensure a more rational, efficient, and intensive use of existing capital.

The Role of Labor

Whereas historical experience with respect to the role of capital is encouraging for today's developing nations, that with respect to the role of labor is much less so. Greater output through a more intensive and efficient utilization of capital carries with it an important corollary; namely, a more efficient and intensive utilization of labor. The length of the working day during the English industrial revolution—twelve or fourteen hours for women and children as well as for adult males—was one of its characteristics most deplored by contemporary reformers as well as by latter-day historians. Nor were such hours peculiar to England; all industrializing nations have experienced them. As recently as the early years of this century the average workday in manufacturing in the United States was just under ten hours, having been reduced from twelve to fourteen in the mid-nineteenth century [17]. In agriculture the average workday was, if anything, even longer.

Such facts are not very agreeable, but they cannot be wished away. I find it surprising that so little attention has been paid to them in discussions of development policies and programs. On the contrary, faced with the demands of organized and/or politically conscious workers, the development policies of most noncommunist nations call for a reduction in the length of the workday or workweek (along with minimum wages and large social insurance benefits on the model of their more affluent neighbors) at the same time that they are striving for a larger output per person. The situation is further aggravated in most developing nations by the existence of substantial pools of unemployed or underemployed workers in both urban and rural areas. This is, to say the least, a very horny dilemma.

I want to make it clear that I am not advocating a return to nine-

teenth-century labor standards. Indeed, there is much evidence, historical and other, that the excessive length of the workday characteristic of early industrialism did not necessarily represent the most efficient use of labor. On the other hand, there are many ways to achieve a more productive use of labor without resorting to crude exploitation. One of the definitional characteristics of an underdeveloped country is the existence of a large, underemployed, low-productivity labor force in agriculture. Evidence from the United States, Japan, and several other countries of the large role of agricultural labor in the formation of agricultural capital is therefore of some relevance [21]. Achieving a more productive use of labor in this fashion will probably require some changes in systems of land tenure—but that is a subject too vast and complicated to be entered into in so brief a disquisition. It is also likely that to secure a more intensive and efficient application of labor in agriculture and elsewhere will require a prolonged process of education, for it is by now obvious that the "propensity to truck and barter" and other aspects of economizing behavior which Adam Smith and his disciples took for granted is not instinctive but culturally conditioned. It is, in short, "learned behavior." The inhabitants of the underdeveloped world have learned to desire and demand the outputs of modern economic systems, but they have not yet learned, or have done so only imperfectly, the much more complicated relationships between inputs, efficiency, and outputs. This dictum applies to entrepreneurs and policy-makers as well as to workers, to private and public systems of incentives and rewards as well as to politically-motivated work stoppages.

Policy planners in developing nations will argue that the higher productivity of labor made possible by modern technology will allow them to enjoy rising per capita incomes without resorting to the "exploitation" of workers that characterized the earlier history of developed nations. That may be so, but modern technology also requires a skilled, literate, and responsible labor force—not to mention intelligent and resourceful entrepreneurs—to utilize it effectively.[1] Until such time as the developing nations can educate and train their labor forces and entrepreneurs to much higher standards than exist today—and for most of them that time is not imminent—they face the bleak prospect of continued low productivity and incomes.

The Role of Education

In recent years the proliferation of studies in the economics of education has convinced most economists of the importance of human cap-

---

[1] It is frequently observed that the technologies developed in industrially advanced nations are not always applicable or advisable for developing nations without substantial modifications. In most cases the modifications involve a smaller scale of enterprise and more intensive use of labor.

ital for the economic growth of advanced industrial nations [11] [12] [13] [27] [28]. As yet, however, there have been few if any studies of the relationship between education and the beginnings of development. Did today's wealthy countries begin the process of modernization with relatively advanced educational systems? Or on the contrary, were they able to create elaborate educational systems with the increments of wealth produced in the process of growth? While there is little doubt that, once growth has begun, the process is one of mutual interaction—knowledge begets wealth, poverty begets ignorance, and vice versa—one would like to know the nature of the relationship at some "time-zero." Anderson and Bowman state, in their analysis of needed research on the role of education in development: "One of the main questions calling for investigation is that of the interplay between the spread of literacy and elementary education in the early stages of the agricultural and industrial revolutions in Western countries and Japan" [3, p. 173]. While I have not made the detailed investigations called for, I would like to present some observations based on general studies of the economic history of the developed nations. (In this as in all other sections of the paper the documentation suffers from the fact that it is being written in one of the developing nations, where library resources are less than overwhelming.)

Britain at the end of the eighteenth century probably had the highest literacy rates in the world at that time. It is estimated that in the 1840's 60 percent of adult males and 40 percent of adult females could read and write; and there is some evidence that literacy was even higher at the end of the eighteenth century than in the 1840's [14]. The special case of Scotland is of exceptional interest. The superiority of the Scottish educational system over that of the English, at every level from primary schools to universities, dated from the Reformation. In the course of the eighteenth century every parish came to have at least one school. Some of the most notable technical and organizational innovations of the industrial revolution originated in Scotland, and in the century from 1750 to 1850 the rate of economic growth in Scotland far outstripped that of England, converting Scotland from a primitive tribal economy into one of the most heavily industrialized of the times [7, Chap. III].

Germany at the end of the eighteenth century was a poor, backward area, in spite of the success of Frederick the Great in building up the military power of Prussia. A more fruitful if less spectacular accomplishment of that ruler was his creation of the nucleus of a broadly-based system of elementary education. In the campaign of reforms stimulated by the humiliating defeat of Prussia in the Napoleonic Wars that system was extended, and in subsequent decades technical

institutions were established, the universities reorganized and revitalized, and new universities were created. By the 1870's German universities were recognized as the world's best, especially in science and engineering, whereas half a century previously they had been either tradition-bound or imitative [6, pp. 52-53]. The important point, however, is that the restructuring and extension of the educational system took place prior to the advent of rapid industrialization.

Similar phenomena occurred in Switzerland, the Netherlands, and the Scandinavian countries, and help to account for their rapid industrialization at the end of the nineteenth century. Limitations of space prevent me from discussing the role of education in U.S. economic development, which in any case should be more familiar to this audience. For the same reason I refrain from commenting on the particular strengths and weaknesses of the French educational system, which I have dealt with elsewhere [6, Chap. III]. One final case should not be overlooked, however. According to R. P. Dore, Japan started "its career of forced-pace modernization with a widespread and well developed tradition of formal institutionalized education" [25, p. 105]. Professor Rosovsky declares that on the eve of the Meiji Restoration of 1868 literacy rates were higher "than in any European country at a comparable stage of development, with the exception of Holland and Prussia " [ibid.]. Although I would like to insert more exceptions, such as Switzerland and the Scandinavian countries, at least on a provisional basis until I see more evidence, the evidence already available is sufficient to justify Rosovsky's assertion that "Japan's stock of education—and human capital—was unusual by international standards" [ibid.]. Moreover, the policies of the Meiji government worked to increase those stocks. According to Lockwood, "no other enterprise of the State paid handsomer dividends to the nation" than its investment in education [18, p. 512].

This lesson for developing nations is unusually clear, I believe. A modern, widespread, and variegated educational system, far from being a luxury which only wealthy nations can afford, is more of a necessary cost which developing nations must incur if their plans and aspirations are to be achieved.

The Role of the State

We arrive, finally, at the most controversial area. What is the optimal role of the state in contemporary developing nations? How does it differ, on the one hand, from the role of the state in those countries that began to industrialize in the nineteenth century on a predominantly private enterprise basis, and, on the other, from those countries—principally the Soviet Union—that have succeeded in building a more

or less modern industrial structure on the basis of comprehensive state control and planning? The subject is certainly highly emotive as well as complicated; it may be foolhardy even to mention it in such a brief and necessarily inconclusive fashion. I wish to emphasize, therefore, that my purpose is not to define the optimal role of the state on the basis of historical experience but merely to indicate how that experience may be relevant even under the vastly different circumstances of present-day developing nations.

The "choice" between comprehensive state control and a less closely regulated economy is not in any meaningful sense of the word a choice at all. For any given country the outcome will be determined by forces that are not amenable to rational decision making. For that reason I restrict my remarks to cases in which rational choice with respect to the nature and extent of state participation in the economy is still possible. I do not mean to imply that Soviet-style economies cannot also utilize rationality in decision making, or benefit from historical experience, nor that other economies cannot learn from their experiences—for example, the inevitability of coercion and regimentation which such systems involve—merely that the subject is sufficiently complicated as it is.

In order to keep ideological interferences to a minimum, let us begin by clearing the ground of old myths and half-truths. It is not true, in the first place, that the classical economists advocated a complete "hands off" policy for the state [22]. Second, it is even farther from the truth to suppose that nineteenth century governments adhered systematically to the doctrines of the economists in that or any other matter. These points are so well established that it may appear gratuitous to repeat them to this audience; I include them for the benefit of the wider audience for whom this paper is ultimately intended.

The exact nature and extent of government participation in the economy varied considerably from country to country and from one period to another within any given country. Britain came the closest to achieving the program of the economists (though one should remember that economists, then as now, were rarely unanimous on any specific policy proposal); but Britain was also the home of Benthamite liberalism, the forerunner of one important collectivist school of thought. At the same time that the government was engaged in dismantling the old clumsy apparatus of restrictive legislation and institutions, it began a whole new series of legislative interferences with individual activity in the name of "the general welfare" or for the protection of the weak and disadvantaged [4].

Outside of Britain governments generally took a far more active role

in the economy, from the decentralized and competitive promotional activities of state and local governments in the United States during the Jacksonian era to the highly centralized and bureaucratized activity of the Russian state under Count Witte in the 1890's. The results, as one might suppose a priori, were highly varied; some enterprises and activities succeeded admirably; others were miserable failures. Thus one cannot generalize completely about the efficacy of government participation from historical experience.

Government participation in the process of production (I leave aside for the moment welfare activities) may take one of two broad forms: (1) it can encourage and stimulate private enterprise, strengthen the operation of free markets, and supplement and correct them when they are too weak or inadequate to produce the desired results; or (2) it can seek to control or eliminate private enterprise and to restrain or replace free markets by government edicts. Generally speaking, government policies in those countries that succeeded in industrializing, including Japan and Russia under Witte, were of the former type, whereas those in countries that notably failed to industrialize, such as Spain and the Ottoman Empire (and Russia for most of the nineteenth century), were of the latter type.

It is my impression that the policies of many developing nations resemble primarily those of the second category mentioned above. In part this is understandable, even if the policies themselves are not wise. Frequently, especially in ex-colonial areas but also in many countries of Latin America, private enterprise is identified with foreign investment, which, whatever its contributions to domestic well-being may or may not have been, is regarded with suspicion. In other cases private enterprise of the traditional type has too often been unenterprising, the ally or offspring of traditional ruling classes and/or corrupt governmental patronage. Whatever the sources of resentment against private enterprise and however earnest and idealistic the desires of government officials of developing nations to better the lot of their fellow citizens under the tutelage of the state, they should consider carefully and dispassionately the possibilities and limitations of direct versus indirect action on the economy. The forces of the market are powerful, even within backward, underdeveloped economies. Harnessed in the interests of society by means of appropriate fiscal and monetary policies and similar instruments, they can contribute importantly to economic growth and development. If, however, the state attempts to overrule or outlaw them, they can produce chaos and stagnation. All too often policy-makers attempt to operate directly on the target variables, where their action is not only inefficacious but down-

right harmful to the cause of economic development; for example, by attempting to curb inflation by means of price controls, while simultaneously expanding the stock of money.

Many of the alleged deficiencies of private enterprise in underdeveloped countries are the result of government activities—frequently contradictory and/or mismanaged—that are justified by references to those very deficiencies. For example, manufacturing industry is encouraged because of its apparently higher productivity than agricultural employment; but the differential in productivity may result at least in part from tariff protection and other favors granted to manufacturing, while agriculture stagnates with inadequate transport, marketing, and credit facilities, and positive disincentives in the form of price controls. On the other hand, industrial entrepreneurs fail to exhibit dynamism because they rest secure in the knowledge that exorbitant tariffs will protect their limited domestic markets but are unable even to try to compete in foreign markets because of tariffs on their raw materials, overvalued currencies, and other disincentives to export.

It is frequently argued that the shortage of capital and entrepreneurs in contemporary underdeveloped economies is the reason why the state must play a larger direct role than was true in the nineteenth century. Such arguments are unsound historically. As I have already indicated, many if not most of today's developing nations have higher rates of capital formation than was the case with the developed nations in their early stages of industrialization. As for the shortage of entrepreneurs, that argument can only be tested after the fact. There was no reason to expect that Scotland in the mid-eighteenth century or that Japan in the 1860's would soon produce some of the world's most resourceful entrepreneurs.

Arguments for a larger direct entrepreneurial role for the state also overlook the fact that decisions and actions must still be taken by men. The kinds of men to be found in the service of the state—and, more importantly, the kinds of incentives that impel them—are not necessarily those most conducive to economic growth. By their very nature, political organizations attract many individuals who have no qualifications for getting a job done except their devotion to the party organization and (all too often) their desire to wield power and dispense patronage. Insofar as there is an actual shortage of trained, skilled, and dedicated manpower, it will be found to be at least as acute in government as in the private sector—and with a far greater potential for harmful effects. Even the ablest of individuals will be rendered ineffective if burdened with too many tasks that cannot be delegated for want of capable subordinates. In the circumstances of most developing nations, Parkinson's Law is no laughing matter; it is

a frightful reality. Many examples, which would be amusing if they were not so tragic, could be cited.

There is little doubt that government officials today are, in general, much more "growth-minded" than their counterparts in the past, but that is not necessarily a valid reason for substituting their energies and decisions for those of private entrepreneurs. It is a common observation that state-sponsored development projects tend toward the grandiose and monumental, frequently of dubious economic value, while small-scale industry and agriculture are often neglected. The most successful historical cases of industrialization, however, show the predominance of small-scale enterprise in the earliest stages of industrialization.[2] If small enterprises are successful, it is because they have met the test of the market, and they will inevitably grow. On the other hand, if they fail, the losses to society are much less than in the case of large government-sponsored enterprises, which for political reasons may not be allowed to appear to have failed.

Let me reemphasize that I am not advocating a laissez faire role for the state. Government has, indeed, many important functions in modern economies, both developed and developing—but the wholesaling of onions and potatoes is not one of them. In addition to maintaining an "exact system of justice"—no small task in itself, and one which is very badly performed in many underdeveloped countries—there are huge tasks ahead in almost all developing nations in providing more adequately for the health and educational needs of their populations, as well as maintaining other essential services which government can do more effectively than private enterprise. These tasks are quite sufficient to employ fully both the fiscal and human resources which most governments can command; until they can operate their railways and postal systems efficiently, they should not consider a petrochemical industry, much less the wholesale distribution of foodstuffs.

In closing I would like to append a yet more personal comment. I am aware that the views expressed in this paper, and especially those of the concluding section, are lacking in scientific rigor. Whether for good or ill, policy making is still more of an art than a science, and discussions of policy necessarily reflect the personal biases of the discussants. Let me make it quite clear, therefore, that my own biases are not based on any doctrinaire ideological grounds whatsoever. On the contrary, as an "old socialist" (or, at any rate, a middle-aged former socialist), whatever ideological proclivities I retain run in the opposite direction. My "conversion," if that is the proper term, is the result of twenty years of intensive study of the historical record and a somewhat shorter but no less intensive exposure to the vagaries of policy

[2] The considerations of fn. 1 apply to this point also.

making in developing nations. History, as I have suggested, is a great teacher—personal history not least of all.

REFERENCES

1. M. Abramovitz, "Resource and Output Trends in the United States since 1870," *A.E.R.,* May, 1956, pp. 5-14.
2. *A.E.R.* "Economic History: Its Contribution to Economic Education, Research and Policy," May, 1965, pp. 86-112.
3. M. J. Bowman, and C. A. Anderson, "The Role of Education in Development" in R. E. Asher, *et al., Development of the Emerging Countries: An Agenda for Research* (Brookings Inst., 1962).
4. J. B. Brebner, "Laissez-faire and State Intervention in 19th Century Britain," *J. of Econ. Hist.,* 1948, pp. 59-73.
5. A. K. Cairncross, "Capital Formation in the Take-off," in *The Economics of Take-off into Sustained Growth,* W. W. Rostow, ed. (London, 1963), pp. 240-60.
6. Rondo Cameron, *France and the Economic Development of Europe, 1800-1914* (Princeton, 1961; rev. ed., Chicago, 1966).
7. ——, *et al., Banking in the Early Stages of Industrialization* (New York, 1967).
8. P. Deane, "Capital Formation in Britain before the Railway Age," *Economic Development and Cultural Change,* No. 3, 1961.
9. P. Deane and W. A. Cole, *British Economic Growth 1688-1959, Trends and Structure* (Cambridge, 1962).
10. P. Deane, *The First Industrial Revolution* (Cambridge, 1965).
11. E. F. Denison, *The Sources of Economic Growth in the United States and the Alternatives before Us* (C.E.D., 1962).
12. S. Fabricant, *Trends in Income and National Product in the U.S.* (N.B.E.R., Occasional Paper No. 63, 1959).
13. F. Harbison and C. Myers, *Education, Manpower and Economic Growth* (New York, 1964).
14. J. F. C. Harrison, Letter to the author, Madison, Wis., Aug. 17, 1966.
15. W. Hoffman, "The Take-off in Germany," in W. W. Rostow, ed., *op. cit.,* pp. 95-118.
16. S. Kuznets, "Notes on the Take-off," in W. W. Rostow, ed., *op. cit.,* pp. 22-43.
17. Stanley Lebergott, *Manpower in Economic Growth: The American Record since 1800* (New York, 1964).
18. W. W. Lockwood, *The Economic Development of Japan* (Princeton, 1954).
19. J. Marczewski, "The Take-off Hypothesis and French Experience," in W. W. Rostow, ed., *op. cit.,* pp. 119-38.
20. R. Nurkse, *Capital Formation in Underdeveloped Countries* (New York, 1962).
21. M. L. Primack, "Farm Capital Formation as a Use of Farm Labor in the United States, 1850-1910," *J. of Econ. Hist.,* 1966, pp. 348-62.
22. L. Robbins, *The Theory of Economic Policy in Classical Political Economy* (London, 1952).
23. P. N. Rosenstein-Rodan, "Problems of Industrialization in Eastern and Southeastern Europe," *Econ. J.,* 1943, pp. 202-11.
24. ——, "International Aid for Underdeveloped Countries," *Rev. of Econ. and Statis.,* May, 1961, pp. 108-36.
25. H. Rosovsky, "Japan's Transition to Modern Economic Growth, 1868-1885," in H. Rosovsky, ed., *Industrialization in Two Systems: Essays in Honor of Alexander Gerschenkron* (New York, 1966).
26. ——, *Capital Formation in Japan, 1868-1940* (New York, 1961).
27. T. W. Schultz, "Investment in Human Capital," *A.E.R.,* 1961, pp. 1-16.
28. ——, "Education and Economic Growth," in H. G. Richey, *Social Forces Influencing American Education* (Chicago, 1966).
29. United Nations, *Yearbook of National Accounts Statistics* (1965).
30. C. Veliz, ed., *Obstacles to Change in Latin America* (London, 1965).

SOURCES OF POSTWAR GROWTH IN NINE
WESTERN COUNTRIES

By EDWARD F. DENISON
Brookings Institution

This session, according to its organizer, was designed to see what individuals studying growth by different approaches have learned. My approach is to try to isolate the sources of observed growth of real national income between selected dates and quantify their contributions. The contribution of each source is expressed as the number of percentage points it contributed to the growth rate.

I used this approach a few years ago to analyze United States growth from 1909 to 1957.[1] My paper today concerns an application of similar techniques to nine countries in the 1950-62 period and 1950-55 and 1955-62 subperiods. J. P. Poullier assisted me throughout this study. The countries covered and their growth rates from 1950 to 1962 are West Germany 7.3, Italy 6.0, France 4.9, Netherlands 4.7, Denmark 3.5, Norway 3.5, United States 3.3, Belgium 3.2, United Kingdom 2.3. In several countries growth rates were quite different in the two subperiods. In some, 1950-55 was quite disturbed by recovery from World War II.

My method of analyzing growth distinguishes broadly between the contribution of changes in inputs and the contribution of changes in output per unit of input. To measure the contribution of labor, capital, and land the change in each of these inputs, subdivided as necessary and possible among components, must first be measured. The growth rate of each input is then multiplied by its share of national income to obtain its contribution to the growth rate of national income.[2] The contribution of all inputs together is the growth rate of total factor input when the separate inputs are combined by use of income share weights. The contribution of output per unit of input is the amount of the growth rate that is not explained by the growth of inputs. To divide the contribution of output per unit of input among its sources required a separate technique for each source. Broadly speaking, I have tried to isolate the contributions made by important changes in resource allocation, by economies of scale, and by a few sources, such as the change in the average age of capital and the balancing of the capital stock in the early years in Germany, that were either not very important or im-

---

[1] Edward F. Denison, *The Sources of Economic Growth in the United States and the Alternatives before Us* (C.E.D., 1962).
[2] A different procedure was adopted for dwellings and international assets.

portant only in a particular country. I have also tried to isolate the effect on growth rates of output per unit of input of differences between the years compared with respect to the pressure of demand upon employed resources and, insofar as it affected farm output, of weather. A couple of adjustments were also made for differences in deflation techniques. I was left with a residual representing the contribution of advances in knowledge, any catching-up of technique, the contribution of all changes not explicitly measured such as in how hard people work, and, of course, noncompensating errors in the growth rates themselves and in the estimates for the sources dealt with explicitly.

Since the Econometric Society is a joint sponsor of this session, I had better state that my methodology makes almost no use of correlation analysis. It is quantitative, however, and thus qualifies as econometric according to the charter of the Society.

To measure labor input I took account, sequentially, of changes in employment, changes in hours worked annually by employed persons, changes in the distribution of man-hours among individuals classified by age and sex, and changes in the composition of weighted man-hours classified by amount of education of the individuals working. I shall say a few words about each.

The United States had the second largest employment increase among the nine countries from 1950 to 1962. Germany had the largest. If there had been no change in the composition of employment, its increase would have contributed 1.5 percentage points to the growth rate in Germany, 0.9 in the United States, 0.8 to 0.4 in the Netherlands, Denmark, the United Kingdom, Italy, and Belgium, and as little as 0.1 in France and Norway.

Annual hours worked by full-time nonfarm wage and salary workers dropped least—almost not at all—in the United States and France. They dropped most, about 13 percent, in Germany, and by intermediate amounts elsewhere. But in the United States much of the increase in employment consisted of women and students voluntarily working part time, most of them very few hours a week. This pulled down the average hours of all workers. In Italy, in contrast, involuntary part-time employment was reduced as job opportunities expanded. Elsewhere changes in part-time employment were minor. I have estimated the net effect on the work done in a year of changes in full-time hours, an assumed partial productivity offset, and changes in the importance of part-time employment. It subtracted 0.2 from the United States growth rate and about as much in five of the eight European countries. The subtraction was larger in Germany. It was negligible in France and the contribution was slightly positive in Italy.

The effect of changes in age-sex composition upon average labor

\

quality was measured by use of hourly earnings weights. The change was most unfavorable in the United States where the share of female man-hours rose most. I estimate it subtracted 0.1 from the United States growth rate. In several countries the contribution was positive —as much as plus 0.1 in France and Italy.

Education is a very important aspect of labor quality. In the United States the education of the labor force has been rising rapidly for several decades. In the 1950-62 period it was rising more in the United States than in any of the European countries and much more than in most of them. Based on a modification of the technique used in my previous study, I estimate that the increased education of the labor force raised the average quality of labor enough to contribute 0.5 percentage points to the United States growth rate from 1950 to 1962. The amount was 0.4 in Belgium and Italy, 0.3 in France and the United Kingdom, 0.2 in the Netherlands and Norway, and only 0.1 in Denmark and Germany. These estimates are based on the use of different weights for the various education levels in the United States and Europe. The United States margin over Europe in the increase in education as such is even greater than these numbers suggest.

When all four aspects are combined, I find that the increase in labor input contributed 1.1 percentage points to the United States growth rate and that only in Germany, where employment increased much more, was the contribution of labor input larger than this.

Capital was divided among four components in this study. The contribution of dwellings to the growth rate can be measured directly by turning to the details of the national accounts to obtain the net value of housing services. The contribution was greatest, 0.25, in the United States. Germany followed with 0.14. The contribution of international assets can also be measured directly. It was 0.05 in the United States, and this amount was exceeded only in the Netherlands. Increases in the stocks of nonresidential structures and equipment and of inventories contributed 0.5 to growth in the United States and this was less than in the European countries except Belgium and the United Kingdom. Germany obtained by far the largest contribution from these sources, 1.4.

All types of capital together contributed 0.8 to the 1950-62 growth rate in the United States and about the same amount in the European countries as a group. The contribution was much larger than this, 1.4, in Germany, moderately higher in the Netherlands and Denmark than in the United States, and appreciably lower in Belgium and the United Kingdom.

The quantity of land was estimated not to have changed significantly, and land therefore made no contribution to growth.

If growth depended only on increases in inputs and if all countries operated under constant returns to scale, 1950-62 growth rates would have been as follows: Germany 2.8, United States 2.0, Netherlands 1.9, Italy 1.7, Denmark 1.6, France and Belgium 1.2, United Kingdom 1.1, and Norway 1.0.

The most pervasive reason that there is so little correspondence between a ranking of countries by growth rates of national income and by growth rates of factor input in this period is to be found, according to my analysis, in the reallocation of resources. I tried to deal with three aspects: the contraction of agriculture, the contraction of non-farm self-employment, and the reduction of barriers to international trade.

All the countries had in 1950, and still have, a larger proportion of total resources, but mainly people, in agriculture than the proportion that would provide a maximum national income. From 1950 to 1962 the percentage of employment devoted to agriculture declined by 30 to 47 percent in each of the nine countries. These are enormous declines to occur in only twelve years. The percentage decline was big everywhere but the importance of agricultural employment, and therefore the effect upon nonfarm employment, varied greatly. In 1950 agricultural employment ranged from 5 percent of total employment in the United Kingdom, 11 in Belgium, and 12 in the United States to 25 to 29 in Germany, Denmark, and France and 43 in Italy.

National income per unit of input is generally much lower in agriculture than in nonfarm industries. In addition, given time for adjustment, the addition of inputs to nonagricultural industries could raise nonagricultural national income, excluding income from dwellings and abroad, about proportionately, whereas agriculture was so overmanned that elimination of labor in agriculture had only a small adverse effect on agricultural output. My estimates of the contribution to the 1950-62 growth rate that was obtained by shifting agricultural resources to nonfarm activities range from less than 0.1 in the United Kingdom and 0.2 in the United States and Belgium to 0.7 in France, 0.8 in Germany, and 1.0 in Italy.

There is overallocation of labor to nonfarm self-employment that is generally similar to the situation in agriculture. Nonfarm proprietors and unpaid family workers represent very different proportions of total nonfarm employment in different countries, and the proportion in each of the nine countries declined from 1950 to 1962. This group, of course, includes a core of professionals, proprietors of substantial establishments, repairmen, and the like, who do well enough operating as proprietors and in no sense are misallocated. But it is not this group who account for differences between places and dates but a fringe

group, usually much larger, who are very inefficiently employed or underemployed. As large numbers were eliminated and transferred to wage and salary employment, the work they did could be replaced by those remaining or by a disproportionately small increase in paid employment in larger establishments. Gains from this transfer, I estimate, ranged from 0.04 percentage points in the United States and United Kingdom to 0.22 to 0.26 in Italy, France, Norway, and the Netherlands.

My estimates of gains from reductions in barriers to international trade, which are very crude, ranged from nothing in the United States and 0.02 in the United Kingdom to 0.15 or 0.16 in Belgium, the Netherlands, Norway and Italy.

The combined contributions made to 1950-62 growth rates by these three aspects of resource reallocation were as follows: United Kingdom 0.1, United States 0.3, Belgium 0.5, Netherlands 0.6, Denmark 0.7, Norway 0.9, France and Germany 1.0, and Italy 1.4. The differences are big.

In the 1955-62 subperiod the contributions of inputs and resource allocation together pretty well established the ranking of the nine countries by growth rates, aside from the effects on output per unit of input of incomparability of the years 1955 and 1962 with respect to demand pressures and weather. In France and perhaps Italy, however, there appears to have been an added ingredient. To discuss it requires reference to the residuals I obtained after deducting from growth rates the contributions made by all sources of growth explicitly estimated. In the United States itself the residual contributed 0.76 in both 1950-55 and 1955-62. After minor adjustments for procedural comparability, my former study yields about the same amount for the period since the 1920's. I think this residual for the United States can be used as a tolerable estimate of the contribution made to growth by advances in knowledge. In 1955-62 I obtained residuals in the narrow range of 0.75 to 0.97 for seven of the nine countries. In addition to the United States they were Belgium, Denmark, the Netherlands, Germany, the United Kingdom, and Norway. Since these are residuals, I do not regard the small differences among these countries as significant. For France, however, I obtained a residual of 1.56 and for Italy of 1.30. The French residual also exceeded that for the United States by about the same amount in 1950-55. I am inclined to believe that the large French residuals indicate that France was obtaining an appreciable contribution to growth from sources I have not isolated—perhaps a catching up of technological and managerial techniques toward the United States level as a result of deliberate efforts and the pressure of increased competition, elimination of redundancy in nonfarm wage and

salary employment, improvement of resource allocation in ways I have not measured, better incentives and harder work, or perhaps something else.

In the 1950-55 period, Germany and, to a lesser extent, Italy were apparently securing large increases in output per unit of input from the elimination of wartime distortions. I tried to estimate the effects of one or two aspects of this, but for the most part my techniques are inadequate to deal with so disturbed a situation and it is reflected in large residuals for Germany and Italy in 1950-55.[3]

My general conclusion thus far is that the ranking of countries with respect to growth rates in the 1950-62 period as a whole was determined by changes in inputs, gains from resource reallocation, some catching up of technique or other unisolated source in France and possibly Italy, and recovery from war distortions in the early postwar years in Germany and Italy.

The size of the differentials among countries in growth rates was, however, widened by economies of scale. To some extent this was simply because gains from economies of scale depend on the rate at which the size of markets increase, so that the more national income rises for other reasons the greater are gains from scale economies. There is, however, a much more important reason, one that I rather despair of describing even sketchily in the time available.

I have estimated the amount by which European growth rates would be reduced if the components of consumption were reweighted by use of United States rather than European prices. For 1950-62 as a whole the amounts range from 0.1 in Belgium, Norway, and the United Kingdom and 0.2 in Denmark and the Netherlands to 0.5 in France, 0.6 in Italy, and 0.9 in Germany. The reason rates are reduced is that, for different products, the ratio of the volume of European to United States consumption is lower and the ratio of European to United States prices is higher the greater is the income elasticity for a product. As European per capita consumption has risen, the increases have been concentrated in products that have high income elasticity and high relative prices in Europe as compared with the United States. The larger the rise in per capita consumption the larger is this dispersion. The greater the increase in per capita consumption and the lower the level of per capita consumption the more a shift to United States weights reduces the growth rate. I used an indirect method of estimate that picks up only this systematic tendency but checked the results against those obtained by direct reweighting.

If you prefer, you can consider the estimates just given as simply a

---

[3] The residual in 1950-55 is also particularly large in the Netherlands and small in Denmark. I have no special explanation.

statistical adjustment required to place all the national income esti-
mates on a common basis with respect to consumption price weights. I
believe, however, that it is predominantly an economies-of-scale phe-
nomenon. As per capita consumption in the fast growing European coun-
tries increased, increases were concentrated in products produced in
particularly small volume and at particularly high unit cost in Europe
as compared with the United States. Techniques of large-scale produc-
tion were already available in the United States and could be adopted
as soon as markets developed to sufficient size.

In addition to sources of growth, I have tried to isolate the sources
of difference in levels of national income per person employed. Mea-
sured in United States prices, national income per person employed in
all the European countries covered except Italy was 58 to 65 percent
of the United States level in 1960. In Italy it was 40 percent.

I now offer two groups of observations that emerge from a compari-
son of sources of growth with sources of difference in income levels.

My first comments concern resource allocation. France and Germa-
ny had a much greater potential to raise national income by increasing
nonfarm wage and salary employment at the expense of agricultural
and self-employment than did the United States, and this provides an
important part of my explanation of why output per unit of input rose
more in France and Germany. Greater overallocation is also one rea-
son the level of income is lower in France and Germany than in the
United States but it is a relatively minor reason. France and Germany
are exhausting this source of growth very rapidly, and its complete
elimination would go only a small way to narrow their income gap with
the United States.

Reallocation of resources also provides much of the explanation of
why French and German growth rates were above the British rate. But
the level of national income per person employed was as high in the
United Kingdom as in France and Germany in 1960 only because the
United Kingdom had much less misallocation. In nonfarm industries
alone output per person was appreciably lower in the United Kingdom
and, indeed, not greatly above Italy. Reduction of misallocation is
opening an income gap between France and Germany, on the one
hand, and the United Kingdom, and closing the gap between the
United Kingdom and Italy.

Agricultural and self-employment depressed the Italian income level
most, and this was a principal reason income in Italy was below the
other European countries. Their eventual contraction to the same pro-
portions as in the other countries would eliminate much of the gap in
income levels. However, education and capital also contribute to the
gap.

My second comment concerns what I can only describe as residual productivity. My estimates imply that if there were no difference in the quantity and quality of labor per worker, the amount of capital and land per worker, the size of markets, the cost of misallocation of resources of the types mentioned, or the pressure of demand upon resources, residual productivity in the European countries as a group, except Italy, would still have been 28 percent below the United States in 1960. Whatever other changes may occur, European income per worker cannot approach the American level unless this gap is greatly reduced. My estimates of growth sources indicate that up to 1962, at least, the gap was not being cut much except in France, and even in France the amount seems small compared to the apparent potential.

The United States in 1925 had already reached the level of national income per person employed that was attained by the European countries except Italy, as a group, in 1960. One striking statistic emerging from my study is that residual productivity was lower in Europe in 1960 than it was in the United States in 1925. The gain in the United States from 1925 to 1960 was mainly due, I believe, to advances in knowledge that accumulated over a 35-year period. But disparities in the availability of knowledge as to how to produce cannot explain the 1960 gap between the United States and Europe. One can hardly avoid speculating as to what is responsible, but I shall not do so today.

I shall close with one final generalization. My study indicates that the continental countries did not obtain higher growth rates than we after 1950 because they were doing more to obtain growth. We were drawing more additional workers, particularly women, into employment than any of the other countries. The education of our labor force increased more. Only we and the French did not reduce the hours of full-time workers significantly. We did the most research and development. We cut down farm employment and fringe nonfarm self-employment as much, proportionately, as the European countries. We did not increase the capital stock as much, in percentage terms, as most of the European countries but this situation is less clear than it may seem, and in any case only Germany obtained much more growth than we from this source. The reasons that European growth rates were higher in the postwar period must be sought primarily in differences in initial conditions.

In conclusion, let me state that I keenly regret that I have been unable to provide here the evidence for my conclusions or to introduce the cautions that are usual and appropriate. A book will remedy these omissions in the near future.[4]

[4] Edward F. Denison with the assistance of Jean-Pierre Poulliet, *Why Growth Rates Differ: Postwar Experience in Nine Western Countries* (to be published by the Brookings Institution, 1967).

DISCUSSION

HENRY ROSOVSKY: It is very difficult to disagree with the "lessons" which Professor Cameron finds in the historical record. History, he tells us, demonstrates the virtues of greater efficiency, hard work, education, and the spirit of private enterprise. But all this seems to me somewhat obvious and much too general. If I were a development planner, a bureaucrat, or a businessman in what is here euphemistically called a "developing nation," I would thank the economic historian with courtesy and sincerity, and then send a cable (probably collect) to Messrs. Chenery and Tinbergen saying: "Please come quickly." I say this, because it seems to me that Professor Cameron's lessons, at least by implication, concentrate more on the attributes of economic backwardness than on ways to achieve economic growth.

If we want to draw operational policy implications from the history of modern economic growth, there are at least three prerequisites:

1. We must have one or more explicit theories of the industrialization process.

2. We must, if the lessons are to be useful, restrict the scope of our observations rather severely.

3. We must also, in order to be honest, accentuate the negative.

Let me now say something about each one of these points as they pertain to Professor Cameron's observations.

I find in Professor Cameron's paper too little attention paid to the changing historical nature of the industrialization process. He tells us that "initial conditions" vary, and that Marx's dictum concerning the "mirror image" is unacceptable. And yet his theory concerning the influence of changing initial conditions is never made explicit, and many of his lessons could lead one to believe that Marx was, after all, correct. (Something which I do not believe.) Take the author's remarks concerning capital. He finds that in advanced countries, during the early phases of industrialization, the proportion of net capital formation to national product was surprisingly low, and that these rates rose only after modern economic growth was well under way. "The big development in getting the growth process under way was a more intensive and efficient use of both the existing stock of capital and of normal increments to that stock." This may have been true when the comparatively rich countries of the world industrialized in the nineteenth century. Since then, however, both the nature of technology and the income levels of countries desiring to industrialize have changed. The more efficient use of existing capital stock may have too little impact in parts of Africa, and "normal increments," whatever that may mean, could be far too slow. Underdeveloped countries must import technology, and much of it is capital intensive. It does little good to remind them that this was less true fifty or seventy-five years ago.

This brings me to the need for restricting the scope of observations. Professor Cameron's developed countries run all the way from the United States— by way of Europe—to Japan. The rest of the world is largely underdeveloped.

Yet the developed and the underdeveloped parts of the world are even less homogeneous than the factor capital. For example, there are—historically and currently—young and relatively empty countries and older and settled countries. In the past, the rates of savings in these countries differed in accordance with this classification, and perhaps the lessons should take these differences into account. (The United States always had a surprisingly high NDCF/NNP ratio. Could this have some useful implications for some countries today?) As things stand, Professor Cameron's conclusions tend to overgeneralize. He says: "The kinds of men to be found in the service of the state—and more importantly the kinds of incentives that impel them—are not necessarily those most conducive to economic growth. . . . It is a common observation that state-sponsored development projects tend towards the grandiose and monumental, frequently of dubious economic value. . . ." If we leave out the qualifying clauses, as we must (the kinds of men to be found anywhere are not necessarily this or that), these statements are far too broad to be useful. The bureaucrats of Meiji Japan and of early nineteenth century Prussia did a lot for economic growth and avoided the monumental. The same may be true of Tito's Yugoslavia. Let me add that I do not think that proves anything one way or the other. My own feeling is that for historical lessons to be helpful we must force ourselves to operate at a much lower level of abstraction. I believe that we could learn a great deal by studying regional patterns of development; i.e., the historical process by which poorer regions of developing countries were drawn into the national market, and the accompanying declines in regional income differentials. How Japan exploited the growth potential inherent in sharp regional variation may contain valuable suggestions for South Korea and Thailand. Beyond that I would not go, but I would remain very satisfied if the assumptions are verified.

My last point has to do with the accentuation of the negative. In our reading of the historical record we all too frequently seem rather too satisfied with ourselves, and we must tend to give the impression that "our way" was the best and most alternatives are quite impossible. But Germany was a leading European aggressor twice in this century; Russia went through a great revolution; Japan tried to conquer much of Asia in the 1930's. These events may suggest some of the disequilibrating influences, or let us say the pathology, of modern economic growth, and they are also part of our history. To me, at least, they illustrate some of the dangers of "growth at any cost," and they deserve full consideration when giving advice to others.

We have listened today to an interesting sermon by a reformed socialist. His present state of mind seems to be rather Weberian and this is as good as any other, because to me, in deducing lessons from the complicated record of history, the most suggestive motto has always been "let a hundred flowers bloom."

CHARLES P. KINDLEBERGER: Today's paper by Mr. Denison is but the tip of a massive iceberg represented by the forthcoming book, *Why Growth Rates Differ,* which he was kind enough to send me in manuscript as background for this comment. It is surely inappropriate for me to review that book before it has appeared, and yet I cannot forebear from saying that just as his

Sources of Growth in the United States and the Alternatives Before Us lifted the discussion of United States growth to a new level of discourse, so the book he has written with the assistance of Jean-Pierre Poullier will mark a new stage in the discussion of European growth and productivity. The profession will be deeply in his debt for the monumental energy, patience, and informed judgment he has brought to the measurement of the sources of European growth from 1950-62 and their comparison with those in the United States. The work is a model of patient and detailed exposition, admirable candor, wide-ranging combing of material, and ingenuity of estimation and assumption. If there is room for doubt as to his conclusions, this is the result not of Mr. Denison and Mr. Poullier's skill being wanting, but of possible limitations in the method.

There is an apocryphal quotation ascribed to Raymond Goldsmith that what cannot be measured does not exist. Mr. Denison comes close to this view at one point where he refers to "noneconomic or unmeasured quality change." On the whole, however, he is content to measure what he can and speculate in restrained fashion about what he cannot. But measurement occasionally distorts, as when quality of education is represented by years of schooling, corrected for changes in school days per year and absenteeism of pupils, without reference to the qualitative distinctions which have long been thought relevant to growth between the education of an Oxford gentleman and that of the French or German scientist at the *Grandes Ecoles* or the *Technische Hochschulen*, respectively. And the need to measure comparatively, over a number of countries, may lose valuable data. In allowing for the gain from reallocation, Denison measures only the movement from the farm and from self-employed nonfarm workers to industry, and chooses to ignore the improvement of efficiency from reallocation within sectors and within plants of varying sizes in manufacturing branches. I believe these effects were of great importance in, say, France. The sort of contrast which the statistical Procrustean bed misses is the well-known difference between British railroads with their large numbers of workers, thought by some observers to be almost half in disguised unemployment, and France where the S.N.C.F. since the war has cut down employment while building up service in a remarkable fashion.

But the basic question must be directed to method. Can the factors contributing to growth be estimated as if the world's production functions are linear and homogeneous so that the contribution of separate inputs can be estimated from their income shares; and are the contributions of individual factors separable? As Mr. Denison's paper states, changes in inputs are first measured in their contribution to growth, with the remaining contribution of changes in output per unit of input ingeniously divided among, mainly, scale, allocation, and knowledge effects, plus a residual which includes such immeasurables as quality of management, effort, differences in competition, etc. (The advance of knowledge is in fact part of the residual, but the extension of the United States estimated residual which is identified as "advance in knowledge" to other countries leaves a still smaller residual for other causes and errors in estimation which do not cancel.)

Are these sources of inputs and of increased efficiency per unit of inputs

separable? Denison recognizes and allows for "interactions" between improved allocation between farm and nonfarm, and from self-employed non-farm to paid employment. He also frequently notes connections between effects measured separately, such as expansion of labor and economies of scale. But he is not prepared to test the hypothesis that economic growth is a systems problem in which the variables have important interrelations. As a student of trade, for example, I am perhaps unduly moved by the statistical result that the reduction of barriers to international trade produced only 8 hundredths of one percentage point, or 2 percent of the total rate of growth of Northwest Europe in 1950 to 1962. Denison deliberately overlooks the contribution of exports and import-substitution to growth on the macroeconomic or demand level, since his is a supply, not a demand analysis, in sharp distinction to those of Maddison, Lamfalussy, Dow, etc. I applaud the attention to supply, and this point does not trouble me. But increased trade is given credit only for a very small improvement in resource allocation from improved comparative advantage, whereas I should have attributed to it some of the improvement due to better allocation between farm and nonfarm sectors, between self-employed nonfarm and other nonfarm, and the unmeasured improvement in efficiency from reallocation within manufacturing and within manufacturing branches.

I suppose my major concern is that I have just finished a study of European postwar growth—largely intuitive rather than quantitative—which applies the W. Arthur Lewis model of growth with unlimited supplies of labor to various countries in Europe and finds it suggestive. In this way of looking at things, the availability of extra labor, either new to the labor force or reallocated from inefficient sectors, holds down wages, sustains profits, and maintains investment at a high level. When the supply of labor ceases to be elastic, increases in demand lead to higher wages, lower profits, reduced investment, a reversal of the balance of payments, and so on. The various elements which Denison studies carefully and quantitatively, one by one, are fitted into a model. There is difficulty in making it fit all cases, and France and Austria pose special difficulties. But unlike Denison's approach, the economy is viewed as a system rather than as an aggregation.

It is of some interest that Mr. Denison's studies end in 1962 just as Europe ran out of the large supplies of labor which had sustained growth from 1950 or so. On my showing, the supergrowth of the 1950's will not be repeated after 1962 or 1963 or 1964. Mr. Denison's analysis may lead to the same conclusion but there is no way of knowing it. I can illustrate my methodological point best, perhaps, by referring to a footnote of Lionel Robbins to the effect that the only impact of the Great Depression on the National Bureau was to make them change their trend lines, in contradistinction to Keynes who used it to find a new insight.

But my major comment is the one with which I started. Mr. Denison and his assistant have given us numbers for analysis and food for thought to occupy us for years to come. It is a superb achievement. Mr. Denison has isolated most and perhaps the most important sources of observed growth in quantitative form. The task is now to reassemble them into a coherent whole.

PROPERTY RIGHTS AND BEHAVIOR

STOCK VERSUS MUTUAL SAVINGS AND LOAN ASSOCIATIONS: SOME EVIDENCE OF DIFFERENCES IN BEHAVIOR*

By ALFRED NICOLS
University of California, Los Angeles

The "mutual" savings and loan association originated as a self-help cooperative in which each saver received one "share" for every $100 of savings up to a maximum of fifty shares.[1] Though ultimate power is today supposed to reside in the shareholders, the latter typically sur-render their proxies to management on opening an account. The conse-quence has been that while the fifty-share maximum has effectively diffused ownership, control has been concentrated in management. Subject to little, if any, pressure from savers, management is a self-per-petuating autocracy. A West Virginia court even held it permissible to pay off a dissenting shareholder who had asked for a list of sharehold-ers in order to conduct a proxy fight. Not surprisingly, one legal au-thority concluded that "a shareholder has a vote until he cares to exer-cise it, and then he ceases to have a vote, if the board so decides."[2] Lenin's retort to Bertrand Russell's complaint that there were no polit-ical parties in the Soviet Union is relevant. "A bourgeois lie. Russia has many parties: the communist party in power; and the rest in pris-on." It is significant that it was not an official of an eastern European "people's democracy" but the executive director of the U.S. Coopera-tive League[3] who rationalized that it was necessary to protect the "beneficial ownership . . . of the depositors—not only now but yester-day and tomorrow."[4] Today's mutual savings and loan association has more in common with the monolithic Soviet dictatorship than with the

[1] Each borrower is permitted one share.
[2] P. James Riordan, Hearings before House of Representatives, Committee on Banking and Currency (Subcommittee on Bank Supervision and Insurance), 85th Cong., *Federal Charter*. Legislation for Mutual Savings Banks, p. 206.
[3] *Ibid.*, p. 296, Testimony, Jerry Voorhis.
[4] It has been observed by Arthur H. Courshon, National League of Insured Savings and Loan Associations: "The current Board has taken the position that the shareholders in a mutual institution are represented by the Board. They are not represented by themselves." *Hearings*, p. 251.

benevolent and democratic cooperative societies of nineteenth-century America.

The only constraint is that management is limited to a wage which must be "reasonable." On the other hand, management has not only been allowed but encouraged to establish privately-owned ancillary corporations which provide services to both the customers and the association: mortgage brokerage, insurance, title, trustee. It is the purpose of this paper to examine both the mechanism by which management converts its power position into personal gain and the evidence relating to the efficiency of the mutual as compared to the stock savings and loan association. Performance is measured in terms of growth in assets, market share changes, expenses, gross operating income, and allocations to reserves.

I. *Market Structure in the Savings and Loan Industry*

The savings and loan industry is similar to Caesar's Gaul in that it is divided into three parts. Associations operating under federal charter, all mutuals, had $56.4 billion in assets at the end of 1963, 52.5 percent of the total assets of the Federal Home Loan Bank system. State chartered mutuals had $29.8 billion in assets, 27.7 percent of the total. State chartered savings and loan associations of the "stock" type have increased by 436 percent or from $4.0 billion in 1955 to $21.2 billion, a rise in market share from 11.0 percent to 19.7 percent. California stock associations constituted 62 percent of total stock assets. Stock associations are restricted to twenty states.

II. *A Model for a Mutual Savings and Loan Association*

At the outset it should be noted that the stress is on the differences between the two types of ownership and control. It is not argued that stock associations are completely run and dominated by their stockholders. Stockholders are interested in earnings; and at times exert pressure on management; and are often successful in disposing of that management. Furthermore, even when stockholders are relatively passive there is always the possibility, however remote, that outsiders will make a raid on the stock of the association.

Management of mutuals are free from such headaches and are proud that their decisions need not show earnings designed to inflate stock prices. They boast that their mutual is not "up for grabs" by the wrong people.[5] Presumably the "right" people are presently in control.

The problem of self-perpetuating management in a mutual is one of conversion. How is managerial power converted into personal wealth? There are choices but each is limited:

[5] *Ibid.* Testimony, J. E. Hoeft, Chairman of the Board and President, Glendale Savings & Loan Association of Glendale, California, pp. 289-90; Adolf A. Berle, pp. 307, 308.

1. Though income may be maximized in the form of compensation or wages, it is not possible to directly convert all the profits. Growth in the size of the savings and loan association in order to increase compensation involves certain disadvantages. First, risk exposure is increased and loss of control where insolvency is a probable result is threatened. Since management only receives the benefits of that growth to a limited extent, it would merely add to its worries. Furthermore, there is an asymmetry between success and failure in a mutual. Management receives only part of the rewards of success and the entire blame for the failure!

2. The management may exercise power in the form of amenities taken on the association's premises; e.g., attractive furnishings, beautiful secretaries, "agreeable" persons reflecting the same ethnic and religious background, association-owned automobiles, lengthy vacations with pay, and large expense accounts. But even amenities are subject to check: (a) the association must work within the constraint that revenues cover all costs if it is to remain solvent; (b) the Internal Revenue Service is quick to identify amenities as payments in lieu of income for which taxes must be paid.

3. The diversion of potential association income into wholly owned ancillary establishments. Management in a mutual savings and loan association is a power unto itself. It determines both output and the prices charged both for the services rendered and the prices paid suppliers.

Consider a disequilibrium situation, where there are only mutual associations in each market and entry is restricted by the regulatory authority. Associations in growth periods would receive prices in excess of operating costs. Where a stock association, operating under normal profit maximizing incentives expands because marginal revenue is in excess of marginal costs, the mutual need not and would not. By not doing so its behavior is similar in nature to a restrictive monopolist. A reluctance to move towards profit equilibrium is expected of all mutuals regardless of the market situation. This relative insensitivity of mutual output to profit establishes a margin greater than that for the profit maximizing equilibrium and is as permanent as is the disequilibrium from which it is derived.

But not only is there less reason for the mutual to offer the equilibrium output, there is also no reason why the market would be permitted to determine prices. Where under conventional profit maximization reduced output would imply an excess demand, or "shortage," which would force price up, the mutual sets price on the basis of expenses. His optimal price is considerably less than one which "clears" the market.

The conversion mechanism rests on the manager's power over out-

put and price. Diversion of income to his affiliates requires permanent disequilibrium or restriction coupled with prices below those which would clear the market. Given the shortage, management sets nominal prices and bargains with individuals over the premium they must pay. The price of a loan includes more than the interest rate and fees but also the purchase of related services. The stock association with no necessity for diversion will produce a larger volume and its prices will be smaller by the value to the borrower from securing the related services in free or more competitive markets.

The mutual management thereby "rations" loans on the basis of private deals with borrowers. For example, he may be more willing to lend to borrowers who purchase insurance from his private insurance company. The borrower gains a loan he would not otherwise receive. The value to the mutual is illustrated: in 1964, had the largest five California mutuals secured only one-third of the insurance on their outstanding loans, each would have picked up an estimated $129,585 additional income. In 1964, 4.13 percent of the assets of federally chartered associations in Los Angeles were held in the form of cash on hand or in commercial banks. For the five largest federals with an average $642 million in assets, each averaged $27 million. With certificates of deposit selling at 4.0 percent, the return would have been $1,000,000. If the management of a large federal decided to forego interest[6] in favor of a demand deposit, it is obvious that a bank would be most appreciative.

The placement of savings and loan association cash flows is not inconsiderable. When "loans repaid" are added to net savings for 1964, the five largest federals would have had about a $148 million flow and the next five $56 million. Commercial banks can create many opportunities for acknowledging appreciation. Lines of credit readily become available to the savings and loan association officials, and may be used to purchase options for speculative land appreciation—an arrangement not permitted officers of the savings and loan associations with their own institution. Or relatives and business associates in other fields, on the reference of the savings and loan official, may secure loans for which they would otherwise not qualify. Whether it be commercial bank deposits or the sale of insurance, escrow, or trustee relationships is immaterial; management is in a position to use the margin between prices and cost in a bargaining arrangement which is similar to a "tie-in" sale.

[6] The New York legislative inquiry of 1905 (The Armstrong Committee) revealed that the large New York insurance companies had maintained unduly large cash balances in banks in which both the officers and their companies were heavily involved. (Joint Committee of the Senate and Assembly, New York, Investigation of Life Insurance Companies, Report, Vol. X, pp. 30-31, 83-84, 132).

The mutual association will not grow as rapidly as the stock association, but will also have less incentive for competitive increases in the dividend rates. Competition at any level is to be avoided because it reduces the margin and thereby destroys the mechanism for diversion. Without the margin, the management has nothing to offer the customer or supplier in return for receipt of a line of credit at the bank or the insurance contract. Mutuality restricts the dimensions of competition in that the management cannot afford to forego the tie-in arrangement. Accordingly, the chances for price competition among mutuals are that much smaller—whatever the initial degree of competition or oligopoly may have been. Greater reliance is placed on ancillary activities.

III. *The Evidence*

The hypothesis derived from the above model is that mutual savings and loan associations, in contrast to stock savings and loan associations, will report a retarded rate of growth associated both with lower gross operating income and lower dividend rates paid to savers. The relevant comparisons are between the California federal and stock associations and the California stock associations as compared to the remainder of the United States (which is dominated by the mutual form). The reason for the two comparisons is that the California federally chartered associations have had to develop some of the attributes of the stock associations in order to survive. Unlike mutuals elsewhere in the nation, they must compete with stock associations. Consequently, the differences between California federals and California stock would not be as large as that between the latter and mutuals outside of California.

Expenses. Emphasis in a mutual is on expenses rather than profit; and is directed to both managerial compensation as well as the indirect and nonpecuniary benefits of the manager's position. When California federal associations are compared to stock associations in 1962, the latter with a new loan volume of $3.8 million was twice the federals with $1.8 million. Yet the stock associations' expenses at $119.3 million were only 43 percent larger than the federals' expenses at $73.8 million (Table 1). On the basis of a variant of Parkinson's Law, expenses are what income permits.

Conventional measurement of expenses on assets or gross operating income is unsatisfactory. Variations in product mix, e.g., the ratio of "loans made" to assets, or the "construction" loan to asset ratio, influence the expense ratio and, therefore, make comparisons misleading. The trouble is that both assets and gross operating income are derived from assets acquired in previous years. Once placed on the books such assets make significantly smaller additions to expenses.

TABLE 1

OPERATING RESULTS: U.S. COMPARED TO CALIFORNIA,
MUTUAL AND STOCK ASSOCIATIONS
1962

	New Loans÷ Average Assets	Expenses ÷Average Assets	Gross Operating Income÷ Average Assets	Fees— Loans Made	Dividends ÷Average Saving Capital	Allocation to Reserve and Sur- plus÷ Average Net Worth
U.S. less California insured members FHLB..........	21.3	1.14	5.55	1.49	4.01	12.6
California federals..	31.8	1.28	6.38	2.32	4.59	18.1
California stock....	41.3	1.31	7.10	3.14	4.84	22.0
			1963			
U.S. less California.	21.2	1.17	5.48	1.29	4.04	9.3
California federals..	34.9	1.36	6.31	2.14	4.72	11.8
California stock....	45.1	1.42	6.60	2.77	4.78	15.4

SOURCE: Federal Home Loan Bank Board, *Combined Financial Statements*, 1961–63, and *Source Books*, 1962–64 (Washington, D.C.).

In order to escape from these difficulties, ratios were abandoned in favor of a multiple regression analysis using actual data with respect to forty-eight of the states and thirty-five[7] of the largest standard metropolitan statistical areas (SMSA's). (See Table 2.) The major explanatory variables were dollar volume of "loans made," "new savings," and the average number of mortgage accounts. Loans made were further subdivided into "construction" loans, "other" loans, and "home purchase."

The regression analysis indicated that if loans made had been doubled in 1962 as compared to the mean for the series, expenses would have been increased by 65.5 percent for the states and 59.1 percent for the SMSA's. When the California and Los Angeles stock data with respect to these variables is related to the regression coefficients, a mutual area with the same product mix would have had expenses of 62.6 percent and 72.6 percent higher than what was actually achieved by the stock associations.[8]

[7] California was excluded because of the influence of the stock associations. Alaska was both minor and isolated. Five cities were excluded because of the dominant influence of stock associations: Los Angeles, Phoenix, San Bernardino, San Jose, and San Francisco.

[8] When allowance is made for estimated greater average size of "loans made" by California institutions, the excess of the mutuals is reduced to 30 percent. But this adjustment is not thought necessary for three reasons: (1) when size of loan is tested as an independent variable, it turned out to be statistically insignificant; (2) it is not necessary to make an adjustment when comparisons are made within the same market area, e.g., California where average size of loans made is roughly identical for both mutuals and stock: (3) an adjust-

TABLE 2

OPERATING COST REGRESSION MODELS

	Constant Term	Loans Made	Number of Mortgages	New Savings	R^2
35 SMSA's 1962.....	−0.33062				.9850
Reg. Coef..........		0.03433	−0.00514	0.01479	
(s.e.)..............		(.00423)	(.01158)	(.00262)	
Part. Cor. Coef.....		0.82429	−.07947	0.71134	
35 SMSA's 1963.....	−1.47961				.9893
Reg. Coef..........		0.03059	0.05646	0.00735	
(s.e.)..............		(.00562)	(.01347)	(.00236)	
Part. Cor. Coef.....		0.69930	0.60139	0.48794	
35 SMSA's 1964.....	−1.46038				.9893
Reg. Coef..........		0.02461	0.06718	0.00812	
(s.e.)..............		(.00648)	(.01725)	(.00384)	
Part. Cor. Coef.....		0.56337	0.57320	0.35510	
48 States 1962 (Mutual only)......	−659				.9888
Reg. Coef..........		9.03551	0.00868	0.01041	
(s.e.)..............		(.00353)	(.00554)	(.00178)	
Part. Cor. Coef.....		0.83491	0.22985	0.66117	

SOURCE: Federal Home Loan Bank Board, *Combined Financial States*, 1962–64, and *Savings and Home Financing: Source Book*, 1963–64.

Trend in Expenses, 1950 to 1964. When the ratio of total operating expenses is averaged for the years 1962-64 and compared to the average for 1950-52, the United States, without California, showed an increase of 434 percent. (See Table 3.) This is compared to an increase in activity as measured by loans made of 285 percent. The number of mortgage accounts increased by 122 percent; and the number of savings accounts, 186 percent. Since the major item in expenses is loan origination, it is significant that the increase in expenses was 52 percent higher than the increase in loans made. California stock expenses increased 1,296 percent, while "loans made" rose 1,363 percent.[9] The ratio of incremental expenses to incremental loans was .95 as compared to 1.52 for the United States without California. Where expenses increased by 434 percent for the U.S., assets increased 457 percent. In contrast, the

ment on the basis of "average size" does not solve problems: it only changes them because average size is a ratio of dollar volume of mortgages to number of mortgages. Mortgages are heterogeneous, differing as to expense involved in placing on the books: riskier loans often entail more work. Therefore, it seemed preferable to eschew tampering with the lid of this Pandora's box. Even in a competitive market, the stock generated twice as much new loan volume with but 43 percent higher expenses. On the basis of 1960 census information indicating differences in size of loans made, California stock data were deflated by 23.8 percent. The difference was only reduced to 41.6 percent.

[9] California stock association data for 1950-52 estimated on the basis of asset and loans made shares.

California stock associations' expenses increased 1,296 percent and assets 1,702 percent.[10]

  Volume of Operations. It has been suggested above that measurement by total assets is misleading since the volume of assets in no way reflects the activity in any one period. To illustrate, an association of $100 million in assets may "originate" $20 million in loans a year, while another association of similar asset size or even less might "originate" $40 million in loans. Obviously, the latter association will have higher expenses.

Limitations of space preclude presentation of data with respect to volume for stock associations as compared to federals in the eleven states where both operate and for which there is data. Suffice it to say,

TABLE 3

TRENDS IN EXPENSES AS RELATED TO OTHER VARIABLES, 1950–64
% Changes

	Expenses	Loans Made	Assets	Number of Mortgages	Number of Savings Accounts
U.S. less California.	434	285	457	122	186
California federals..	499	390	609	277	336
California stocks...	1,296	1,363	1,702	554	1,039

SOURCE: Federal Home Loan Bank Board, *Combined Financial Statements*, 1950–64.

in the years 1960-62, the stock associations' ratio of loans made to average mortgages was higher in nine states for each of the years. In addition, this greater ratio of loans made also brought the stock associations' higher gross operating and net income as related to average assets. (See Table 1.)

  Market Share Changes. For California, the market share of the federally chartered associations fell from 60.6 percent of total assets in 1950 to 33.7 percent in 1964. This is consistent with the hypothesis that the mutual associations will show a retarded rate of growth.

  Changes in Net Worth and Assets for California Associations, 1953-62 (Existing December 31, 1953): Federal Compared to Stock. For associations in existence at the end of 1953 (see Table 4), the federals increased assets by 323 percent. The stock associations increased 640 percent. But more significant is the increase in net worth since this

[10] The comparison here repeats the possible shortcoming indicated in note 8: the relatively higher increase in percentage of loans made by California stock associations could be associated with a greater increase in average size of loans made. On the basis of estimated changes in average mortgage size 1950-64, U.S. (less California) was 82.1 percent of California. California stock loans made were deflated with the result that the increase in expenses came to 115.8 percent of the increase in new loans. Notwithstanding the adjustment, the U.S. increase was 31.5 percent more than this deflated estimate.

represents interest-free earning assets. The federal associations' net worth increased $319.6 million, or 327 percent. The stock associations increased $607.9 million, or 649 percent.

Nepotism. It is expected that since nepotism is both a nonpecuniary reward and a tax-free form of inheritance, it is larger in mutuals. Nepotism is defined as a relationship of the chief executive officer to other officers. The relationship is simply the same last name and obviously understates the degree of nepotism because it fails to include in-laws or relatives with different surnames. In 1955, the stock associations reported 27 incidents for 146 members, or a ratio of 18 percent; the federally chartered, 22 incidents for 66 members, or 33 percent.[11]

TABLE 4

CHANGES IN NET WORTH AND ASSETS FOR CALIFORNIA ASSOCIATIONS EXISTING
ON DECEMBER 31, 1953: FEDERAL COMPARED TO STOCK*
(000's)

	1953	1962	Absolute Change	% Change
Net worth† federal.......	97,676	417,243	319,567	327.
stock.........	93,633	701,567	607,934	649.
Total assets federal.......	1,440,404	6,097,108	4,656,704	323.
stock.........	1,234,591	9,141,204	7,906,613	640.

* *1963 Roster of Members*, Federal Home Loan Bank of San Francisco; California Savings and Loan Commissioner, *69th Annual Report*.
† *Combined Financial Statements*, 1953 and 1962, Federal Home Loan and Bank Board.

For 1963, the stock showed 15 percent and the federally chartered, 43 percent.

Management Changes, 1950-64.[12] Change in management is defined as a change in the chairman, president, or listed manager, or any combination thereof. Expressing the management changes as a percentage of the number of associations, the median value reported for the stock associations was 18.9 percent; for the federal associations, 5.9 percent.

Competition for Savings. The California stock associations made greater use of advertising media than did the federals: 59 compared to 12 federals.[13] They also stressed interest rate where the federals referred to "ratio of reserve to savings." When federals are compared to stock within the same states, the ratio of dividends paid to average savings capital was higher for the stock for the years 1960-62. Stock associations were also more effective merchandisers, as revealed in their development and aggressive use of premiums and bonuses. In

[11] Federal Home Loan Bank of San Francisco, *Roster of Members*, 1955 and 1963.
[12] *Ibid.*, 1950-64.
[13] Data derived from *New York Times*, 1955-63, first Sunday edition of each quarter; and similar method used for Los Angeles *Times*.

short, the federal have, on the average, competed only most reluctantly.[14]

IV. *Significance of the Findings*

If mutuality is characterized by public benovolence and stock ownership private cupidity, it might be better if the nation opted for the latter. Mutuality is a euphemism which has effectively shielded powerful monopolistic forces from the public scrutiny. Furthermore, if California experience is any guide, the mutual is an institution which can survive only at the cost of continued governmental restriction of competition; i.e., through achieving a "sacred cow" status similar to agriculture.

The remedy for the excessive margins of mutuality is obvious. If mutuals were converted into stock associations, not only would profit provide a basis for rational decision making, but the necessity for maintenance of nonequilibrium margins would vanish. Removal of the profit constraint on management eliminates the pressure to "second best" choices such as nepotism, infrequent management changes, and tie-in "deals"—all of which increase the price to borrowers while reducing the interest paid to savers. The diversion mechanism is not only restrictive, it is inefficient.

[14] Not only have the stock associations paid higher rates, they initiated the increases last summer when Home Savings and Loan of Los Angeles, the nation's largest, went to 5¼, 0.40 more than the Federal Home Loan Bank set for member associations requesting advances.

TOWARD A THEORY OF PROPERTY RIGHTS

By HAROLD DEMSETZ
University of Chicago

When a transaction is concluded in the marketplace, two bundles of property rights are exchanged. A bundle of rights often attaches to a physical commodity or service, but it is the value of the rights that determines the value of what is exchanged. Questions addressed to the emergence and mix of the components of the bundle of rights are prior to those commonly asked by economists. Economists usually take the bundle of property rights as a datum and ask for an explanation of the forces determining the price and the number of units of a good to which these rights attach.

In this paper, I seek to fashion some of the elements of an economic theory of property rights. The paper is organized into three parts. The first part discusses briefly the concept and role of property rights in social systems. The second part offers some guidance for investigating the emergence of property rights. The third part sets forth some principles relevant to the coalescing of property rights into particular bundles and to the determination of the ownership structure that will be associated with these bundles.

The Concept and Role of Property Rights

In the world of Robinson Crusoe property rights play no role. Property rights are an instrument of society and derive their significance from the fact that they help a man form those expectations which he can reasonably hold in his dealings with others. These expectations find expression in the laws, customs, and mores of a society. An owner of property rights possesses the consent of fellowmen to allow him to act in particular ways. An owner expects the community to prevent others from interfering with his actions, provided that these actions are not prohibited in the specifications of his rights.

It is important to note that property rights convey the right to benefit or harm oneself or others. Harming a competitor by producing superior products may be permitted, while shooting him may not. A man may be permitted to benefit himself by shooting an intruder but be prohibited from selling below a price floor. It is clear, then, that property rights specify how persons may be benefited and harmed, and, therefore, who must pay whom to modify the actions taken by persons. The recognition of this leads easily to the close relationship between property rights and externalities.

Externality is an ambiguous concept. For the purposes of this paper, the concept includes external costs, external benefits, and pecuniary as well as nonpecuniary externalities. No harmful or beneficial effect is external to the world. Some person or persons always suffer or enjoy these effects. What converts a harmful or beneficial effect into an externality is that the cost of bringing the effect to bear on the decisions of one or more of the interacting persons is too high to make it worthwhile, and this is what the term shall mean here. "Internalizing" such effects refers to a process, usually a change in property rights, that enables these effects to bear (in greater degree) on all interacting persons.

A primary function of property rights is that of guiding incentives to achieve a greater internalization of externalities. Every cost and benefit associated with social interdependencies is a potential externality. One condition is necessary to make costs and benefits externalities. The cost of a transaction in the rights between the parties (internalization) must exceed the gains from internalization. In general, transacting cost can be large relative to gains because of "natural" difficulties in trading or they can be large because of legal reasons. In a lawful society the prohibition of voluntary negotiations makes the cost of transacting infinite. Some costs and benefits are not taken into account by users of resources whenever externalities exist, but allowing transactions increases the degree to which internalization takes place. For example, it might be thought that a firm which uses slave labor will not recognize all the costs of its activities, since it can have its slave labor by paying subsistence wages only. This will not be true if negotiations are permitted, for the slaves can offer to the firm a payment for their freedom based on the expected return to them of being free men. The cost of slavery can thus be internalized in the calculations of the firm. The transition from serf to free man in feudal Europe is an example of this process.

Perhaps one of the most significant cases of externalities is the extensive use of the military draft. The taxpayer benefits by not paying the full cost of staffing the armed services. The costs which he escapes are the additional sums that would be needed to acquire men voluntarily for the services or those sums that would be offered as payment by draftees to taxpayers in order to be exempted. With either voluntary recruitment, the "buy-him-in" system, or with a "let-him-buy-his-way-out" system, the full cost of recruitment would be brought to bear on taxpayers. It has always seemed incredible to me that so many economists can recognize an externality when they see smoke but not when they see the draft. The familiar smoke example is one in which negotiation costs may be too high (because of the large number of interact-

ing parties) to make it worthwhile to internalize all the effects of smoke. The draft is an externality caused by forbidding negotiation.

The role of property rights in the internalization of externalities can be made clear within the context of the above examples. A law which establishes the right of a person to his freedom would necessitate a payment on the part of a firm or of the taxpayer sufficient to cover the cost of using that person's labor if his services are to be obtained. The costs of labor thus become internalized in the firm's or taxpayer's decisions. Alternatively, a law which gives the firm or the taxpayer clear title to slave labor would necessitate that the slaveowners take into account the sums that slaves are willing to pay for their freedom. These costs thus become internalized in decisions although wealth is distributed differently in the two cases. All that is needed for internalization in either case is ownership which includes the right of sale. It is the prohibition of a property right adjustment, the prohibition of the establishment of an ownership title that can thenceforth be exchanged, that precludes the internalization of external costs and benefits.

There are two striking implications of this process that are true in a world of zero transaction costs. The output mix that results when the exchange of property rights is allowed is efficient and the mix is independent of who is assigned ownership (except that different wealth distributions may result in different demands).[1] For example, the efficient mix of civilians and military will result from transferable ownership no matter whether taxpayers must hire military volunteers or whether draftees must pay taxpayers to be excused from service. For taxpayers will hire only those military (under the "buy-him-in" property right system) who would not pay to be exempted (under the "let-him-buy-his-way-out" system). The highest bidder under the "let-him-buy-his-way-out" property right system would be precisely the last to volunteer under a "buy-him-in" system.[2]

We will refer back to some of these points later. But for now,

---

[1] These implications are derived by R. H. Coase, "The Problem of Social Cost," *J. of Law and Econ.*, Oct., 1960, pp. 1-44.

[2] If the demand for civilian life is unaffected by wealth redistribution, the assertion made is correct as it stands. However, when a change is made from a "buy-him-in" system to a "let-him-buy-his-way-out" system, the resulting redistribution of wealth away from draftees may significantly affect their demand for civilian life; the validity of the assertion then requires a compensating wealth change. A compensating wealth change will not be required in the ordinary case of profit maximizing firms. Consider the farmer-rancher example mentioned by Coase. Society may give the farmer the right to grow corn unmolested by cattle or it may give the rancher the right to allow his cattle to stray. Contrary to the Coase example, let us suppose that if the farmer is given the right, he just breaks even; i.e., with the right to be compensated for corn damage, the farmer's land is marginal. If the right is transferred to the rancher, the farmer, not enjoying any economic rent, will not have the wherewithal to pay the rancher to reduce the number of head of cattle raised. In this case, however, it will be profitable for the rancher to buy the farm, thus merging cattle raising with farming. His self-interest will then lead him to take account of the effect of cattle on corn.

enough groundwork has been laid to facilitate the discussion of the next two parts of this paper.

The Emergence of Property Rights

If the main allocative function of property rights is the internalization of beneficial and harmful effects, then the emergence of property rights can be understood best by their association with the emergence of new or different beneficial and harmful effects.

Changes in knowledge result in changes in production functions, market values, and aspirations. New techniques, new ways of doing the same things, and doing new things—all invoke harmful and beneficial effects to which society has not been accustomed. It is my thesis in this part of the paper that the emergence of new property rights takes place in response to the desires of the interacting persons for adjustment to new benefit-cost possibilities.

The thesis can be restated in a slightly different fashion: property rights develop to internalize externalities when the gains of internalization become larger than the cost of internalization. Increased internalization, in the main, results from changes in economic values, changes which stem from the development of new technology and the opening of new markets, changes to which old property rights are poorly attuned. A proper interpretation of this assertion requires that account be taken of a community's preferences for private ownership. Some communities will have less well-developed private ownership systems and more highly developed state ownership systems. But, given a community's tastes in this regard, the emergence of new private or state-owned property rights will be in response to changes in technology and relative prices.

I do not mean to assert or to deny that the adjustments in property rights which take place need be the result of a conscious endeavor to cope with new externality problems. These adjustments have arisen in Western societies largely as a result of gradual changes in social mores and in common law precedents. At each step of this adjustment process, it is unlikely that externalities per se were consciously related to the issue being resolved. These legal and moral experiments may be hit-and-miss procedures to some extent but in a society that weights the achievement of efficiency heavily, their viability in the long run will depend on how well they modify behavior to accommodate to the externalities associated with important changes in technology or market values.

A rigorous test of this assertion will require extensive and detailed empirical work. A broad range of examples can be cited that are consistent with it: the development of air rights, renters' rights, rules for

liability in automobile accidents, etc. In this part of the discussion, I shall present one group of such examples in some detail. They deal with the development of private property rights in land among American Indians. These examples are broad ranging and come fairly close to what can be called convincing evidence in the field of anthropology.

The question of private ownership of land among aboriginals has held a fascination for anthropologists. It has been one of the intellectual battlegrounds in the attempt to assess the "true nature" of man unconstrained by the "artificialities" of civilization. In the process of carrying on this debate, information has been uncovered that bears directly on the thesis with which we are now concerned. What appears to be accepted as a classic treatment and a high point of this debate is Eleanor Leacock's memoir on *The Montagnes "Hunting Territory" and the Fur Trade.*[3] Leacock's research followed that of Frank G. Speck[4] who had discovered that the Indians of the Labrador Peninsula had a long-established tradition of property in land. This finding was at odds with what was known about the Indians of the American Southwest and it prompted Leacock's study of the Montagnes who inhabited large regions around Quebec.

Leacock clearly established the fact that a close relationship existed, both historically and geographically, between the development of private rights in land and the development of the commercial fur trade. The factual basis of this correlation has gone unchallenged. However, to my knowledge, no theory relating privacy of land to the fur trade has yet been articulated. The factual material uncovered by Speck and Leacock fits the thesis of this paper well, and in doing so, it reveals clearly the role played by property right adjustments in taking account of what economists have often cited as an example of an externality—the overhunting of game.

Because of the lack of control over hunting by others, it is in no person's interest to invest in increasing or maintaing the stock of game. Overly intensive hunting takes place. Thus a successful hunt is viewed as imposing external costs on subsequent hunters—costs that are not taken into account fully in the determination of the extent of hunting and of animal husbandry.

Before the fur trade became established, hunting was carried on primarily for purposes of food and the relatively few furs that were required for the hunter's family. The externality was clearly present. Hunting could be practiced freely and was carried on without assessing its impact on other hunters. But these external effects were of such

[3] Eleanor Leacock, *American Anthropologist* (American Anthropological Asso.), Vol. 56, No. 5, Part 2, Memoir No. 78.
[4] Cf., Frank G. Speck, "The Basis of American Indian Ownership of Land," *Old Penn Weekly Rev.* (Univ. of Pennsylvania), Jan. 16, 1915, pp. 491-95.

small significance that it did not pay for anyone to take them into account. There did not exist anything resembling private ownership in land. And in the *Jesuit Relations,* particularly Le Jeune's record of the winter he spent with the Montagnes in 1633-34 and in the brief account given by Father Druilletes in 1647-48, Leacock finds no evidence of private land holdings. Both accounts indicate a socioeconomic organization in which private rights to land are not well developed.

We may safely surmise that the advent of the fur trade had two immediate consequences. First, the value of furs to the Indians was increased considerably. Second, and as a result, the scale of hunting activity rose sharply. Both consequences must have increased considerably the importance of the externalities associated with free hunting. The property right system began to change, and it changed specifically in the direction required to take account of the economic effects made important by the fur trade. The geographical or distributional evidence collected by Leacock indicates an unmistakable correlation between early centers of fur trade and the oldest and most complete development of the private hunting territory.

> By the beginning of the eighteenth century, we begin to have clear evidence that territorial hunting and trapping arrangements by individual families were developing in the area around Quebec. . . . The earliest references to such arrangements in this region indicates a purely temporary allotment of hunting territories. They [Algonkians and Iroquois] divide themselves into several bands in order to hunt more efficiently. It was their custom . . . to appropriate pieces of land about two leagues square for each group to hunt exclusively. Ownership of beaver houses, however, had already become established, and when discovered, they were marked. A starving Indian could kill and eat another's beaver if he left the fur and the tail.[5]

The next step toward the hunting territory was probably a seasonal allotment system. An anonymous account written in 1723 states that the "principle of the Indians is to mark off the hunting ground selected by them by blazing the trees with their crests so that they may never encroach on each other. . . . By the middle of the century these allotted territories were relatively stabilized."[6]

The principle that associates property right changes with the emergence of new and reevaluation of old harmful and beneficial effects suggests in this instance that the fur trade made it economic to encourage the husbanding of fur-bearing animals. Husbanding requires the ability to prevent poaching and this, in turn, suggests that socioeconomic changes in property in hunting land will take place. The chain of reasoning is consistent with the evidence cited above. Is it inconsistent with the absence of similar rights in property among the southwestern Indians?

Two factors suggest that the thesis is consistent with the absence of

[5] Eleanor Leacock, *op. cit.,* p. 15.
[6] Eleanor Leacock, *op. cit.,* p. 15.

similar rights among the Indians of the southwestern plains. The first of these is that there were no plains animals of commercial importance comparable to the fur-bearing animals of the forest, at least not until cattle arrived with Europeans. The second factor is that animals of the plains are primarily grazing species whose habit is to wander over wide tracts of land. The value of establishing boundaries to private hunting territories is thus reduced by the relatively high cost of preventing the animals from moving to adjacent parcels. Hence both the value and cost of establishing private hunting lands in the Southwest are such that we would expect little development along these lines. The externality was just not worth taking into account.

The lands of the Labrador Peninsula shelter forest animals whose habits are considerably different from those of the plains. Forest animals confine their territories to relatively small areas, so that the cost of internalizing the effects of husbanding these animals is considerably reduced. This reduced cost, together with the higher commercial value of fur-bearing forest animals, made it productive to establish private hunting lands. Frank G. Speck finds that family proprietorship among the Indians of the Peninsula included retaliation against trespass. Animal resources were husbanded. Sometimes conservation practices were carried on extensively. Family hunting territories were divided into quarters. Each year the family hunted in a different quarter in rotation, leaving a tract in the center as a sort of bank, not to be hunted over unless forced to do so by a shortage in the regular tract.

To conclude our excursion into the phenomenon of private rights in land among the American Indians, we note one further piece of corroborating evidence. Among the Indians of the Northwest, highly developed private family rights to hunting lands had also emerged—rights which went so far as to include inheritance. Here again we find that forest animals predominate and that the West Coast was frequently visited by sailing schooners whose primary purpose was trading in furs.[7]

[7] The thesis is consistent with the development of other types of private rights. Among wandering primitive peoples the cost of policing property is relatively low for highly portable objects. The owning family can protect such objects while carrying on its daily activities. If these objects are also very useful, property rights should appear frequently, so as to internalize the benefits and costs of their use. It is generally true among most primitive communities that weapons and household utensils, such as pottery, are regarded as private property. Both types of articles are portable and both require an investment of time to produce. Among agriculturally-oriented peoples, because of the relative fixity of their location, portability has a smaller role to play in the determination of property. The distinction is most clearly seen by comparing property in land among the most primitive of these societies, where crop rotation and simple fertilization techniques are unknown, or where land fertility is extremely poor, with property in land among primitive peoples who are more knowledgeable in these matters or who possess very superior land. Once a crop is grown by the more primitive agricultural societies, it is necessary for them to abandon the land for several years to restore productivity. Property rights in land among such people would require policing cost for several years during which no sizable output is obtained. Since to provide for

The Coalescence and Ownership of Property Rights

I have argued that property rights arise when it becomes economic for those affected by externalities to internalize benefits and costs. But I have not yet examined the forces which will govern the particular form of right ownership. Several idealized forms of ownership must be distinguished at the outset. These are communal ownership, private ownership, and state ownership.

By communal ownership, I shall mean a right which can be exercised by all members of the community. Frequently the rights to till and to hunt the land have been communally owned. The right to walk a city sidewalk is communally owned. Communal ownership means that the community denies to the state or to individual citizens the right to interfere with any person's exercise of communally-owned rights. Private ownership implies that the community recognizes the right of the owner to exclude others from exercising the owner's private rights. State ownership implies that the state may exclude anyone from the use of a right as long as the state follows accepted political procedures for determining who may not use state-owned property. I shall not examine in detail the alternative of state ownership. The object of the analysis which follows is to discern some broad principles governing the development of property rights in communities oriented to private property.

It will be best to begin by considering a particularly useful example that focuses our attention on the problem of land ownership. Suppose that land is communally owned. Every person has the right to hunt, till, or mine the land. This form of ownership fails to concentrate the cost associated with any person's exercise of his communal right on that person. If a person seeks to maximize the value of his communal rights, he will tend to overhunt and overwork the land because some of the costs of his doing so are borne by others. The stock of game and the richness of the soil will be diminished too quickly. It is conceivable that those who own these rights, i.e., every member of the community, can agree to curtail the rate at which they work the lands if negotiating and policing costs are zero. Each can agree to abridge his rights. It is obvious that the costs of reaching such an agreement will not be zero. What is not obvious is just how large these costs may be.

Negotiating costs will be large because it is difficult for many per-

---

sustenance these people must move to new land, a property right to be of value to them must be associated with a portable object. Among these people it is common to find property rights to the crops, which, after harvest, are portable, but not to the land. The more advanced agriculturally based primitive societies are able to remain with particular land for longer periods, and here we generally observe property rights to the land as well as to the crops.

sons to reach a mutually satisfactory agreement, especially when each hold-out has the right to work the land as fast as he pleases. But, even if an agreement among all can be reached, we must yet take account of the costs of policing the agreement, and these may be large, also. After such an agreement is reached, no one will privately own the right to work the land; all can work the land but at an agreed upon shorter workweek. Negotiating costs are increased even further because it is not possible under this system to bring the full expected benefits and expected costs of future generations to bear on current users.

If a single person owns land, he will attempt to maximize its present value by taking into account alternative future time streams of benefits and costs and selecting that one which he believes will maximize the present value of his privately-owned land rights. We all know that this means that he will attempt to take into account the supply and demand conditions that he thinks will exist after his death. It is very difficult to see how the existing communal owners can reach an agreement that takes account of these costs.

In effect, an owner of a private right to use land acts as a broker whose wealth depends on how well he takes into account the competing claims of the present and the future. But with communal rights there is no broker, and the claims of the present generation will be given an uneconomically large weight in determining the intensity with which the land is worked. Future generations might desire to pay present generations enough to change the present intensity of land usage. But they have no living agent to place their claims on the market. Under a communal property system, should a living person pay others to reduce the rate at which they work the land, he would not gain anything of value for his efforts. Communal property means that future generations must speak for themselves. No one has yet estimated the costs of carrying on such a conversation.

The land ownership example confronts us immediately with a great disadvantage of communal property. The effects of a person's activities on his neighbors and on subsequent generations will not be taken into account fully. Communal property results in great externalities. The full costs of the activities of an owner of a communal property right are not borne directly by him, nor can they be called to his attention easily by the willingness of others to pay him an appropriate sum. Communal property rules out a "pay-to-use-the-property" system and high negotiation and policing costs make ineffective a "pay-him-not-to-use-the-property" system.

The state, the courts, or the leaders of the community could attempt to internalize the external costs resulting from communal property by allowing private parcels owned by small groups of person with similar

interests. The logical groups in terms of similar interests, are, of course, the family and the individual. Continuing with our use of the land ownership example, let us initially distribute private titles to land randomly among existing individuals and, further, let the extent of land included in each title be randomly determined.

The resulting private ownership of land will internalize many of the external costs associated with communal ownership, for now an owner, by virtue of his power to exclude others, can generally count on realizing the rewards associated with husbanding the game and increasing the fertility of his land. This concentration of benefits and costs on owners creates incentives to utilize resources more efficiently.

But we have yet to contend with externalities. Under the communal property system the maximization of the value of communal property rights will take place without regard to many costs, because the owner of a communal right cannot exclude others from enjoying the fruits of his efforts and because negotiation costs are too high for all to agree jointly on optimal behavior. The development of private rights permits the owner to economize on the use of those resources from which he has the right to exclude others. Much internalization is accomplished in this way. But the owner of private rights to one parcel does not himself own the rights to the parcel of another private sector. Since he cannot exclude others from their private rights to land, he has no direct incentive (in the absence of negotiations) to economize in the use of his land in a way that takes into account the effects he produces on the land rights of others. If he constructs a dam on his land, he has no direct incentive to take into account the lower water levels produced on his neighbor's land.

This is exactly the same kind of externality that we encountered with communal property rights, but it is present to a lesser degree. Whereas no one had an incentive to store water on any land under the communal system, private owners now can take into account directly those benefits and costs to their land that accompany water storage. But the effects on the land of others will not be taken into account directly.

The partial concentration of benefits and costs that accompany private ownership is only part of the advantage this system offers. The other part, and perhaps the most important, has escaped our notice. The cost of negotiating over the remaining externalities will be reduced greatly. Communal property rights allow anyone to use the land. Under this system it becomes necessary for all to reach an agreement on land use. But the externalities that accompany private ownership of property do not affect all owners, and, generally speaking, it will be necessary for only a few to reach an agreement that takes these effects into account. The cost of negotiating an internalization of these effects

is thereby reduced considerably. The point is important enough to elucidate.

Suppose an owner of a communal land right, in the process of plowing a parcel of land, observes a second communal owner constructing a dam on adjacent land. The farmer prefers to have the stream as it is, and so he asks the engineer to stop his construction. The engineer says, "Pay me to stop." The farmer replies, "I will be happy to pay you, but what can you guarantee in return?" The engineer answers, "I can guarantee you that I will not continue constructing the dam, but I cannot guarantee that another engineer will not take up the task because this is communal property; I have no right to exclude him." What would be a simple negotiation between two persons under a private property arrangement turns out to be a rather complex negotiation between the farmer and everyone else. This is the basic explanation, I believe, for the preponderance of single rather than multiple owners of property. Indeed, an increase in the number of owners is an increase in the communality of property and leads, generally, to an increase in the cost of internalizing.

The reduction in negotiating cost that accompanies the private right to exclude others allows most externalities to be internalized at rather low cost. Those that are not are associated with activities that generate external effects impinging upon many people. The soot from smoke affects many homeowners, none of whom is willing to pay enough to the factory to get its owner to reduce smoke output. All homeowners together might be willing to pay enough, but the cost of their getting together may be enough to discourage effective market bargaining. The negotiating problem is compounded even more if the smoke comes not from a single smoke stack but from an industrial district. In such cases, it may be too costly to internalize effects through the marketplace.

Returning to our land ownership paradigm, we recall that land was distributed in randomly sized parcels to randomly selected owners. These owners now negotiate among themselves to internalize any remaining externalities. Two market options are open to the negotiators. The first is simply to try to reach a contractual agreement among owners that directly deals with the external effects at issue. The second option is for some owners to buy out others, thus changing the parcel size owned. Which option is selected will depend on which is cheaper. We have here a standard economic problem of optimal scale. If there exist constant returns to scale in the ownership of different sized parcels, it will be largely a matter of indifference between outright purchase and contractual agreement if only a single, easy-to-police, contractual agreement will internalize the externality. But, if there are several externalities, so that several such contracts will need to be negotiated, or

if the contractual agreements should be difficult to police, then
purchase will be the preferred course of action.

The greater are diseconomies of scale to land ownership t
will contractual arrangement be used by the interacting neig
settle these differences. Negotiating and policing costs will
pared to costs that depend on the scale of ownership, and p;
land will tend to be owned in sizes which minimize the sum
costs.[8]

The interplay of scale economies, negotiating cost, externali
the modification of property rights can be seen in the most
"exception" to the assertion that ownership tends to be an in
affair: the publicly-held corporation. I assume that significan
mies of scale in the operation of large corporations is a fact a
that large requirements for equity capital can be satisfied mor
ly by acquiring the capital from many purchasers of equit)
While economies of scale in operating these enterprises exist
mies of scale in the provision of capital do not. Hence, it
desirable for many "owners" to form a joint-stock company.

But if all owners participate in each decision that needs to
by such a company, the scale economies of operating the comp
be overcome quickly by high negotiating cost. Hence a deleg
authority for most decisions takes place and, for most of these
management group becomes the *de facto* owners. Effective ov
i.e., effective control of property, is thus legally concentrated
agement's hands. This is the first legal modification, and it tal
in recognition of the high negotiating costs that would other
tain.

The structure of ownership, however, creates some ex
difficulties under the law of partnership. If the corporation sh(
partnership law commits each shareholder to meet the debts of
poration up to the limits of his financial ability. Thus, mana;
facto ownership can have considerable external effects on shar(
Should property rights remain unmodified, this externality wou
it exceedingly difficult for entrepreneurs to acquire equity capi
wealthy individuals. (Although these individuals have rec(
reimbursements from other shareholders, litigation costs will l
A second legal modification, limited liability, has taken place t
the effect of this externality.[9] *De facto* management owner:
limited liability combine to minimize the overall cost of operati
enterprises. Shareholders are essentially lenders of equity cap
not owners, although they do participate in such infrequent dec

[8] Compare this with the similar rationale given by R. H. Coase to explain the fi
Nature of the Firm," *Economica*, New Series, 1937, pp. 386–405.

[9] Henry G. Manne discusses this point in a forthcoming book about the Am
porate system.

those involving mergers. What shareholders really own are their shares and not the corporation. Ownership in the sense of control again becomes a largely individual affair. The shareholders own their shares, and the president of the corporation and possibly a few other top executives control the corporation.

To further ease the impact of management decisions on shareholders, that is, to minimize the impact of externalities under this ownership form, a further legal modification of rights is required. Unlike partnership law, a shareholder may sell his interest without first obtaining the permission of fellow shareholders or without dissolving the corporation. It thus becomes easy for him to get out if his preferences and those of the management are no longer in harmony. This "escape hatch" is extremely important and has given rise to the organized trading of securities. The increase in harmony between managers and shareholders brought about by exchange and by competing managerial groups helps to minimize the external effects associated with the corporate ownership structure. Finally, limited liability considerably reduces the cost of exchanging shares by making it unnecessary for a purchaser of shares to examine in great detail the liabilities of the corporation and the assets of other shareholders; these liabilities can adversely affect a purchaser only up to the extent of the price per share.

The dual tendencies for ownership to rest with individuals and for the extent of an individual's ownership to accord with the minimization of all costs is clear in the land ownership paradigm. The applicability of this paradigm has been extended to the corporation. But it may not be clear yet how widely applicable this paradigm is. Consider the problems of copyright and patents. If a new idea is freely appropriable by all, if there exist communal rights to new ideas, incentives for developing such ideas will be lacking. The benefits derivable from these ideas will not be concentrated on their originators. If we extend some degree of private rights to the originators, these ideas will come forth at a more rapid pace. But the existence of the private rights does not mean that their effects on the property of others will be directly taken into account. A new idea makes an old one obsolete and another old one more valuable. These effects will not be directly taken into account, but they can be called to the attention of the originator of the new idea through market negotiations. All problems of externalities are closely analogous to those which arise in the land ownership example. The relevant variables are identical.

What I have suggested in this paper is an approach to problems in property rights. But it is more than that. It is also a different way of viewing traditional problems. An elaboration of this approach will, I hope, illuminate a great number of social-economic problems.

ON THE DISTINCTION BETWEEN PUBLIC
AND PRIVATE GOODS*

By OTTO A. DAVIS, *Carnegie Institute of Technology*
and ANDREW B. WHINSTON, *Purdue University*

I. *Introduction*

Some years ago, in a now classic series of articles [13] [14] [15], Professor Paul A. Samuelson made an admittedly polar distinction between public and private goods. Briefly, private consumption goods, like bread, must be parceled out among persons with one man getting a loaf more if another gets a loaf less. Thus, if x_{ij} represents the ith person's consumption of the jth private good and X_j represents the quantity available, then $\sum_i x_{ij} = X_j$. A public consumption good, on the other hand, differs in that one man's consumption does not diminish the quantity available for another. Thus, if y_{ik} represents the ith person's consumption of the kth public good and Y_k represents the quantity available, $y_{ik} = Y_k$ for all men. Prime examples are supposed to be outdoor circuses, national defense, and, some may have thought, radio waves and TV signals.

Recently, responding to Minasian's argument [10] that market allocation (via subscription) of television might be a more desirable arrangement, Professor Samuelson [16] claimed that the possession by television of the characteristics of a public good did not constitute evidence either for or against a market arrangement via subscription. There is no doubt but that Samuelson's claim (or admission?) is correct. Aside from television and radio signals, there are many examples of goods which are allocated in the market place and yet which exhibit various degrees of "Samuelsonian publicness." The products of the researcher's efforts, the outpuring of the jukebox and the record player, undercapacity indoor (why be restricted to the outdoors?) circuses and performances of movies and various forms of entertainment, the unhurried and uncrowded viewing of old masters which are owned by a private art collector, etc.—all exhibit aspects of publicness. Olsen [12] claims that the concept can be applied to all large organizations.[1]

[1] But see Wagner's review [18] of Olsen's theme.

When one is faced with such examples, it is only natural to wonder if it is possible to determine in any objective fashion whether a good should be governmentally or privately produced. Samuelson's classic articles seem to provide few clues to an answer. Yet, the problem appears to merit consideration.

II. *Institutional Arrangements and Samuelson's Model*

It is important to understand that, despite the title, Samuelson's model is not directly concerned with public expenditures. Rather, it is a model of market failure. What is shown is that in a system incorporating his "public" consumption goods, the market "fails" in the sense that the necessary conditions for the attainment of Pareto optimality are not automatically satisfied. One should observe two facts here. First, despite the name given to the public goods, the failure of the pricing mechanism to satisfy the necessary conditions for Pareto optimality does not constitute a prima facie case for public ownership, or even public regulation, of the relevant facilities for either production or distribution. Second, the institutional arrangements under which a good, even a public consumption good, is distributed in a market perform an important role in determining the characteristics of the performance of the pricing mechanism.

Recall the definition of Samuelson's public good. One of the implicit notions is that whatever is produced is available to all. Obviously, even if the appropriate maximization problem yields the necessary optimality conditions with the associated Lagrange multipliers, it is difficult to even imagine real prices performing the role indicated by the multipliers. Behaviorally, the problem appears to be much more complicated and rich than is indicated by the simple mathematics. At issue is the fact that paying the price does not give one control of the good. Instead, the act of paying the price is somehow separated from the act of consumption. Consider the following example in order to clarify the behavioral implications of this point.

Imagine that the government decided to alter the institutional arrangements or organization of the market for bread, which is certainly a private good. Suppose that the government decided to separate the act of paying from the act of obtaining the bread. Accordingly, imagine that payments for the desired quantity of bread were made by consumers to the government in the early morning at the "revenue center." Later in the day, bread could be obtained at the "distribution center." However, suppose that there is no communication between the revenue and distribution centers so that the acts of paying and obtaining are truly separated. The bread is distributed on a first come, first served basis. One might observe that this institutional arrangement would certainly

affect the functioning of the market mechanism. Consumers coul
counted upon to reveal their preferences at the revenue cent
would the distribution center fare much better since consume
simply take the bread they wanted, if it happened to be avail
matter what their payment at the revenue center. Price could
form its traditional functions. The institutional arrangemeni
market failure.

Observe that, at least under one interpretation, there are
similarities between the above example and the Samuelsonian r
public consumption goods. By definition, a public good in Sam
model is available to all consumers no matter whether a c(
chooses to pay or not. From the standpoint of an individual con:
payment could not give one control over the public good in the
deciding the precise quantity to be consumed. The act of p
separated from the act of consuming. It is not surprising that the
fails. For a consumer actually to make a payment would be
more or less than an act of pure charity.

III. *Some Conditions Needed for the Operation of a Mark*

Since he was interested in the problem of attaining an effi
Pareto optimal allocation, Samuelson did not discuss the possi
making a market for public goods operate much as if it were al
private goods. Nevertheless, such a possibility is often availab
alternative to the collective provision of the public good if the a;
ate institutional arrangements can be accomplished. Accordingl}
and informal discussion of some of the conditions which see1
needed if a market is to operate in a reasonable manner appea
merited.

It is obvious that the first requirement is a provision for some
ownership or property rights. Without ownership, at least in son
there would seem to be no basis for exchange. Of course, there a
plex issues involved in the definition of property rights. T
definition can affect the manner in which exchange takes place
way in which a market operates.[2] The definition of property ri{
sometimes determine whether externalities are allowed to exis
market. Zoning ordinances, for example, can be interpreted as a
to remove the influence of externalities in the urban property r
It might be noted that in a sense a good may be characterizec
control an owner may exercise, so that the same piece of prope:
effect a different good once the zoning ordinance is changed.

The second requirement, which is related to the issues of ow

---

[2] This basic point seems to have been first pointed out by Coase [4] [5]. See also
and Stubblebine [3] and Davis and Whinston [7].
[3] See Davis [6].

and property rights, concerns "control" over the good or service. The act of buying or paying for the good or service must be related to the use or consumption of it. Even in the case of public good, purchase can sometimes be made to give some control and establish a connection between the act of paying and consumption as the example of television signals with scramblers illustrates.

The third requirement, which is related to both of the previous ones, concerns the possibility of exclusion. While perfect exclusion need not be required for markets to function reasonably well, as numerous examples of the presence of externalities amply illustrates, it is easily seen that exclusion does provide basic motivation for exchange.[4]

IV. *The Market Allocation of Samuelsonian Public Goods*

In part, the question of whether the above requirements can be satisfied for some particular good depends upon the technological characteristics of the good. In order to illustrate this point, consider the case of television. In the age of electronic scramblers, exclusion is possible. The signals can be "owned" and their services sold. Thus it is possible to establish a direct relationship between the consumption of the service and the payment for it. The above requirements for a market can be satisfied. Since the issue does not seem to have been considered fully, it appears appropriate to discuss briefly the operation of such a market on the basis of an excessively simple model.

Consider a society composed of people deciding upon viewing a pay television program. For simplicity, let the capital or fixed costs of producing the program be ignored. Assume that there is no advertising. Consider the following definitions:

$$x_i = \begin{cases} 0 \text{ if the } i\text{th person does not watch the program} \\ 1 \text{ if that person does watch the program} \end{cases}$$

$$y = \begin{cases} 0 \text{ if the program is not produced} \\ 1 \text{ if the program is produced} \end{cases}$$

$u_i(x_i)$: The value or "utility" to the ith person from watching the program.

$g(y)$: A function indicating the amount of some homogeneous (for simplicity) resource used to produce the program.

M: An arbitrarily large number.

K: The "minimal" amount of the homogeneous resource used to produce the program; i.e., $9(1) = K$.

Given these definitions, and assuming that the objective is to attain a

---

[4] Musgrave [11] emphasizes the notion of exclusion.

Pareto optimum in television viewing, consider the following vector maximization problem:

(1)

$$\text{(1.1)}\quad \max\ [u_1(x_1),\ u_2(x_2),\ \cdots,\ u_n(x_n)]$$

subject to

$$\text{(1.2)}\quad \sum_i x_i - My \le 0$$

$$\text{(1.3)}\quad g(y) \le K$$

$$\text{(1.4)}\quad x_i = 0,\ 1;\ y = 0,\ 1 \qquad i = 1,\ \cdots,\ n$$

It is convenient to replace the vector (1.1). Following the procedure introduced in [8], let α_i represent the reciprocal of the ith individual's marginal utility of income. Define $d_i = u_i(1)$ and $c_i = \alpha_i d_i$. Assuming for simplicity that the α_i are not affected by the decision to watch or not watch the program, one can replace the vector (1.1) by

$$\text{(2)}\qquad \sum_i \alpha_i u_i(x_i) = \sum_i \alpha_i d_i x_i = \sum_i c_i x_i$$

and maximize (2) subject to the same constraints so that the problem is reduced to one of integer maximization. Note that the use of M in (1.2) means that once $y = 1$ the constraint is never binding. The program must be available if someone watches it.

Let p represent the price of the homogeneous resource and assume that p is given by a perfect market. Define $k = p \cdot g(1)$ so that $p \cdot g(y) = ky$. Problem (1) can now be replaced by the following equivalent problem:

(3)

$$\max \sum_i c_i x_i - ky$$

subject to

$$\sum_i x_i - My \le 0$$

and the solution to (3) is obvious. If $\sum_i c_i \ge k$ for $c_i \ge 0$, then the program should be presented and all x_i corresponding to $c_i \ge 0$ should be set equal to one. If there is an i such that $c_i < 0$, then that $x_i = 0$. All persons who enjoy the program should be able to view it.

Consider whether the above solution could be achieved by a non-discriminatory pricing device. Let λ denote the price of being able to view the program. Obviously, if no one who wishes to see the program is to be excluded, then the best that the television network can do is to set

$$\text{(4)}\qquad \lambda = \min_i c_i \ge 0$$

so that the total revenue to the network is $\lambda \sum_i x_i$ where the sum is over all i whose $c_i \geq 0$. Two issues are relevant here. First, $\lambda \sum x_i < k$ is a likely outcome so that a subsidy may be required to pay for the costs of the program. The fact that total revenue is less than variable costs (recall that fixed costs are ignored here) does not necessarily indicate that the program is unwarranted. Second, there seems to be no feasible way for the network to determine the appropriate value of λ as indicated in (4). Obviously, the network cannot ask potential viewers to reveal the value of their c_i. One might think of starting with a zero price and raising it by successive increments until some viewer turned off his set; but surely viewers would catch on to this little game and bluff by turning off their set at any positive price so that it would be driven back down to zero. Therefore, it does not appear that pricing, with or without subsidization, can result in the achievement of the solution to (3).

Since the presence of market failure does not imply the existence of a superior nonmarket alternative, consider the "second-best" type of problem where the television program is priced, there is no subsidy and costs must be covered. It seems obvious in these circumstances that the "best of the second best" possibilities is one which allows costs to be covered exactly. In other words, values of both λ and the x_i must be determined such that

$$(5) \qquad\qquad \lambda \sum_i x_i - k = 0$$

and the other conditions are also satisfied. It is not always true that such a solution exists or that it is unique. Assuming existence, presumably one could find such a solution by starting with an extremely low price and then raising it by increments until (5) was satisfied. Note that the existence of (5) removes the incentive for bluffing. In other words, in this second-best type of allocation by pricing, there is no problem of individuals emitting false signals or not revealing true preferences. Observe, however, that this second-best solution does impose costs of exclusion upon the society. This cost is given by $\sum c_i$ where the sum is taken over those c_i satisfying the relation $0 \leq c_i \leq \lambda$ where here λ represents the particular value of λ which was selected to satisfy (5).

Note that no motivational premise, such as profit maximization, was introduced as the basis for the selection of the second-best constraint (5). However, it can be argued that in an expanded model where alternate networks compete for viewers of their programs, the very competition will cause the "no profit" condition (5) to be satisfied so that this analysis does provide a basis (at least as a first approximation) for the assessment of the costs of the private provision of television services by the pricing mechanism.

Observe again that the particular pricing arrangement analyzed depended upon the technological feasibility of scrambling (an scrambling) devices. However, even if these devices were not ava: market allocations are still possible if certain institutional actioi taken. Law might require that television sets be made in such a m that a special device is required to tune into any given channel purchase of one of these special devices would give the owner of t the right to view the specified channel as often as desired. Such : rangement would not markedly differ from the practice of many p: swimming clubs, for example, of selling a season pass instead of cha for each trip to the pool. Similarly, channels might be assigned 1 networks who might "rent" the right to make sets capable of t into the specified channels to the various manufacturers. Obvii various arrangements are possible. Each possibility has its own oi ing characteristics and corresponding social cost.

The point here is that technological considerations can dete: partially which institutional arrangements are feasible. The feasi of various institutional arrangements certainly has an influence i determination of which particular one is to be selected as the mo: propriate. Since technology changes over time, one can expect th stitutions should be modified accordingly.

V. *Institutional Choice for Samuelsonian Public Goods*

The recent exchange between Minasian and Samuelson, whicl concerned with the issue of subscription television, illustrates thi that the optimality conditions are only of limited value in makii institutional choice.[5] Samuelson [16] emphasizes that there is nc sumption that the particular arrangement which does satisfy the r sary conditions is one which produces a Pareto optimal allocation. cient conditions must also be considered. Even more important must determine whether there are feasible institutional arranger which could result in the satisfaction of both sets of conditions. obviously a fact that there is nothing inherent in the derivation of ε necessary or sufficient conditions for a Pareto optimum which suξ that feasible arrangements for satisfying these conditions exist. interesting to note that Samuelson stressed this very point in the fi his series of papers on this topic.[6]

The consequence of a lack of feasible institutional arrangemen: satisfying both necessary and sufficient conditions for Pareto optin in systems which include public goods is the realization that institut choice involves the comparison of alternative arrangements whic

[5] See Minasian [10] and Samuelson [16].
[6] See Samuelson [13], pp. 388–89.

necessarily nonoptimal in the sense of Pareto. The problem is not to choose from the Pareto optimum positions that one considered ethically desirable. Rather, the problem is to choose from a feasible set of institutional arrangements that particular one which gives the most suitable or "best" allocation of the good under consideration.

It is obvious that in choosing between alternative institutional arrangements the actual operating characteristics of these various arrangements are factors of great importance. As Buchanan has argued so eloquently, it makes little sense to compare operating characteristics of unobtainably ideal arrangements in making an institutional choice.[7] Consequently, if one is choosing, for example, between the alternative institutional arrangements of subscription television, zero pricing and financing by advertising, or zero pricing without advertising with governmental provision of the service, it is not sufficient to argue that subscription television should be ruled out because a nonzero price violates the necessary condition for efficiency, that financing by advertising results in programs appealing to mass taste so that cultural and educational values are overlooked, and that these reasons indicate governmental provision. One cannot simply assume (and be correct) that governmental provision will be "ideal" simply because the other two arrangements are not. One must consider the actual operating characteristics of all alternative arrangements. The problem of institutional choice often involves the comparison of alternative problems of second best which characterize the available possibilities.

For a Samuelsonian public good, there are certain costs associated with private provision and exclusion. When the requirements for a market are satisfied, at the expense of those costs imposed upon the system by exclusion and the operation of a pricing system, one obtains those advantages provided only by markets.[8] Information concerning desires is provided and can be incorporated into decisions concerning supply, quality of the good, etc. It is worthwhile noting that the relative intensity of desires can be at least partially revealed in even an imperfect market system while this kind of information is much more difficult to obtain under any other arrangement. Nevertheless, the costs of exclusion are not to be taken lightly.

Governmental provision of a public good at, say, a zero price, on the other hand, also has problems associated with it. How are the decision-makers to decide, for example, the quantity which should be provided? One can give the easy answer to this question by saying that the political process makes the decision. However, such an answer only evades

[7] See Buchanan [1].

[8] The point that markets are costly to operate has been emphasized by Coase [5]. Demsetz [9] also makes this point but with a greater emphasis upon the costs associated with policing the property rights.

the issue. How does the policitican get information which would lead to the proper decision? One must realize that politicians make decisions at least in part upon the basis of political costs which might be approximated in terms of votes. There does not appear to be any reason to suspect that votes are superior indicators of desires in a system where there are many issues and few elections, so that a vote may enjoy alternative interpretations.

There are obvious difficulties in analyzing those problems of second best where the institutional arrangement is governmental provision, or governmental regulation, of the good under consideration. One does not yet know the appropriate behavioral characteristics which should determine the form of the second-best constraints in the models. What are the behavioral rules which guide the decisions and the actions of governmental agencies? It appears that this question must be answered in a satisfactory and useful manner before there will be a very reasonable basis for analyses capable of adequately determining whether given "public" goods should be governmentally produced and distributed or left to the private sector, with or without regulation.

Finally, it should be noted that ethical considerations cannot completely be ignored in institutional choice. Strotz [17] emphasizes the fact that public goods have important distributional implications. Similarly, alternative institutional arrangements for the allocation of public goods have distributional implications.

VI. *The Allocation of Non-Samuelsonian Public Goods*

It has long been recognized that Samuelson's admittedly polar definition of public goods omits much of governmental activity. It can be argued that another class of goods which should be either governmentally provided or regulated is characterized by extreme decreasing costs. This class of goods involves a high, fixed, and negligible marginal cost coupled with the presence of a capacity constraint. Roads and bridges are prime examples. Leaving aside the question of whether these goods "should" be governmentally produced and distributed, and also leaving the question of how the government actually would do it, let us consider the problem of how such a good "should" be produced and its services allocated over time. One needs to have the benchmark of ideal performance in order to either access or influence actual performance.

Consider a good (say a bridge or a road) which must be constructed in an initial period and whose services must be allocated in that period and also T periods in the future. Supposedly this good becomes obsolete and disintegrates at the end of the Tth period. One incurs the high fixed cost in the initial period and only negligible costs thereafter. Two interrelated questions appear. One must choose the capacity (or decide "how much

to consume") in the initial period and also determine an allocation over time.

In order to answer the above questions, let us consider a very simple model in which there are only two goods. One good (say bread) is produced and consumed during each period. It is an ordinary private good. The other good is the one discussed above. In our system it is often (but not necessarily always) a governmental good and for convenience will be termed such here. Consider the following definitions:

$x^i{}_1(t)$: The amount of the private good consumed by the ith person during period t.

$x^i{}_2(t)$: The amount of the governmental good consumed by the ith person during period t.

$y_1(t)$: The quantity of the private good produced and available for consumption during period t. For simplicity, assume no storage or carryover between periods.

$y_2(1)$: The quantity of the governmental good produced during the initial period and available for consumption during each of the T periods.

$g(y_1(t))$: An implicit "production possibility" function relating available resources (assumed to be one) to $y_1(t)$ during periods $t = 2, \cdots, T$.

$H(y_1(1), y_2(1))$: An implicit "production possibility" function relating available resources (assumed to be one) to the quantities $y_1(1)$, $y_2(1)$ of the two goods which are produced during the initial period.

$U_t{}^i$: The assumed concave utility function of the ith individual during period t.

It is convenient to simplify the problem by ignoring side issues and assuming that there are no externalities in the system. Consequently, it is assumed that in the initial period the productions of the two goods are not functionally interrelated so that H is separable and can be written in the form

(6) $$H(y_1(1), y_2(1)) = h_1(y_1(1)) + h_2(y_2(1))$$

and that the decisions of the initial period do not affect the production possibilities of the following periods so that g can be given as indicated. It is also assumed that utility depends only upon consumption in the given period so that there are no functional interdependencies over time. In addition, it is convenient to assume that utility functions can be written in the separable form

(7) $$U_t^i(x_1^i(t), x_2^i(t)) = u_{1t}^i(x_1^i(t)) + u_{2t}^i(x_2^i(t))$$

so that there is no interaction. The utility from bread does not influence the utility from crossing the bridge and vice versa.

Since interest here is centered upon deriving the conditions for Pareto optimality, it is obvious that the problem is one of vector maximization. It is convenient, however, to apply the Kuhn-Tucker equivalence theorem and follow the procedure outlined in [8] in order to state the criterion function in a more appropriate form. Accordingly, and as before, let α_i represent the (unspecified but assumed positive) reciprocal of the ith consumer's marginal utility of income. The maximization problem can be written as follows:

$$(8.1) \quad \max \sum_i \sum_t \alpha_i u_{1t}^i(x_1^i(t)) + \sum_i \sum_t \alpha_i u_{2t}^i(x_2^i(t))$$

subject to

(8)

$$(8.2) \quad \sum_i x_1^i(t) \leq y_1(t) \qquad t = 1, \cdots, T$$

$$(8.3) \quad \sum_i x_2^i(t) \leq y_2(1) \qquad t = 1, \cdots, T$$

$$(8.4) \quad h_1(y_1(1)) + h_2(y_2(1)) \leq 0$$

$$(8.5) \quad g(y_1(t)) \leq 0 \qquad t = 2, \cdots, T$$

$$(8.6) \quad x_1^i(t) \geq 0, \; x_2^i(t) \geq 0, \; y_1(t) \geq 0, \; y_2(1) \geq 0 \qquad \begin{matrix} t = 1, \cdots, T \\ i = 1, \cdots, n \end{matrix}$$

Note that (8.1) is the criterion (social welfare) function which is stated in a form useful for determining the conditions for Pareto optimality. Constraint (8.2) states that no more of the private good can be consumed than is available in any given period. Constraint (8.3) indicates that no more of the governmental good can be consumed in any period than is made available in the first period. Note that the quantity available of the governmental good is measured in terms of capacity. Constraints (8.4) and (8.5) indicate that no more of the goods can be produced than is allowed by the available resources. Finally, (8.5) indicates the nonnegativity conditions.

It is obvious from the very structure of problem (8) that the usual and familiar conditions apply for the private good. Accordingly, these Pareto conditions are not presented here. Competitive markets are fully capable of satisfying these conditions and performing the desirable allocation. Attention is centered on the conditions for the production and allocation of the governmental good. Let $\lambda(t)$ represent the multiplier associated with constraint (8.3). Recall that this multiplier can be interpreted as a shadow price. Then the conditions on the demand side are

$$(9.1) \quad \alpha_i \frac{\partial u_{2t}^i}{\partial x_2^i(t)} - \lambda(t) \begin{Bmatrix} \leq \\ = \end{Bmatrix} 0 \quad \text{if} \quad x_2^i(t) \begin{Bmatrix} = \\ > \end{Bmatrix} 0, \quad \begin{matrix} i = 1, \cdots, n \\ t = 1, \cdots, T \end{matrix}$$

which are also the conventional and familiar Pareto conditions. If the governmental good is chosen, then an individual should consume that amount in each period which equates his weighted marginal utility to the price of the good during the period. Let β represent the multiplier associated with (8.4). Then on the supply side the condition is

$$(9.2) \qquad \sum_t \lambda(t) - \beta \frac{\partial h_2}{\partial y_2(1)} \begin{Bmatrix} \leq \\ = \end{Bmatrix} 0 \quad \text{if} \quad y_2(1) \begin{Bmatrix} = \\ > \end{Bmatrix} 0$$

which has an interesting interpretation. Note that if one were to determine the quantity (or capacity) of the governmental good to be made available according to a profit maximization criterion with the prices taken as given, then one would consider the problem

$$(10) \qquad \max_{y_2(1) \geq 0} \left\{ \sum_t \lambda(t) y_2(1) - \beta h_2(y_2(1)) \right\}$$

and the conditions for the solution to this problem are given by (9.2). Note that condition (9.2) can be interpreted as saying that if the governmental good is to be supplied, then that quantity (capacity) should be chosen in the initial period which will equate the sum of the prices to the marginal cost of supplying the selected quantity. The solution is that given by profit maximization.

Under the usual assumptions, the conditions derived from (8) are both necessary and sufficient for Pareto optimality. Observe that, in regard to the governmental good, one obtains as a solution to (8) a vector of consumption quantities $(x_2^i(1), \cdots, x_2^i(T))$ for each i, a number $y_2(1)$ which is the quantity made available in each period (the capacity), and a vector of prices or charges $(\lambda(1), \cdots, \lambda(T))$. This solution is Pareto optimal. Note, however, that it cannot be decentralized period by period.

There are some interesting aspects to the solution to this problem which have important implications for the planning for the provision of the class of goods (such as bridges and roads) under consideration. Note that if demand is at all variable, it is likely that the charge $\lambda(t)$ which rations the available supply $y_2(1)$ is likely to be zero for some periods. From the point of view of planning, the zero charge does not imply no exclusion. It merely means that a person does not have to pay to obtain the right to use the bridge, but he must still obtain that right. Also note the implication of (9.2). The capacity (quantity supplied) should not be selected to be so large that the corresponding constraint (8.3) never becomes binding. The constraint must become binding during some

periods so that the charge becomes positive. If demand is such that the charge can never become positive, the facility is not justified.[9] The implication for planning for this class of goods is that one should make a forecast of possible usage over the life of the facility. This forecast should be conditional upon the use of the appropriate charges during each period. Given that revenues can cover costs, one selects the appropriate capacity accordingly.[10] Once the facility is constructed, actual prices (as opposed to the forecasted ones) should be simply adjusted by the conventional rule to ration the available supply.

In terms of practical planning of bridges and roads, where it is usually assumed to be not feasible to use direct charges as a rationing device, the implications of the above analysis would seem to indicate that facilities should not be constructed so as to eliminate congestion during all periods. The constraint (8.3) must be binding during some periods in order to have a rational allocation of resources. Since congestion costs serve (imperfectly) as prices, the construction of facilities which never became congested during their useful life should serve as a clear indication of a misallocation of resources.[11] Rational action should not eliminate congestion but merely obtain a (here undefined) appropriate amount of it.

VII. *Concluding Comments*

This paper has not included a discussion of whether the so-called "governmental goods" should really be produced and distributed by the public sector. The issues here are much the same as those discussed for the so-called "public goods." All that is indicated in the previous section is a benchmark analysis against which the behaviors of private, regulated or public actors could be measured. It is true that analytic models of public actors are needed before institutional choices can be made on anything approaching a reasonable basis. Nevertheless, it does appear to be a shame that public goods are called "public."

[9] Note, however, that supply is assumed to be infinitely divisible.

[10] Recall, however, the convenient assumption of no externalities. If there are externalities, then one has to make an adjustment in this rule to take them into account.

[11] One must make an exception here for the very real problem caused by indivisibility in urral areas. Recall the convenient assumption that the size of the facility was continuously divisible.

REFERENCES

1. James M. Buchanan, "Politics, Policy, and the Pigovian Margins," *Economica*, Feb., 1962, pp. 17–28.
2. ——, "An Economic Theory of Clubs," *Economica*, Feb., 1965, pp. 1–14.
3. James M. Buchanan and W. Craig Stubblebine, "Externality," *Economica*, Nov., 1962, pp. 371–84.
4. Ronald H. Coase, "The Federal Communications Commission," *J. of Law and Econ.*, Oct., 1959, pp. 1–40.
5. ——, "The Problem of Social Cost," *J. of Law and Econ.*, Oct., 1960, pp. 1–44.
6. Otto A. Davis, "Economic Elements in Municipal Zoning Decisions," *Land Econ.*, Nov., 1963, pp. 375–86.

DISCUSSION

STANLEY ENGERMAN: Paul David's paper represents the most thorough study in the debate, set off by the income estimates of Martin and their detailed criticism by Kuznets, concerning American economic growth between 1800 and 1840. It is based both on new time series not previously utilized and upon the resourceful application of available data to clarify certain key points. The new time series are Lebergott's reworking of labor force data, total and by industrial distribution (usefully revised for certain years by David), and the Towne-Rasmussen series on nineteenth-century agricultural output. David has also prepared 1790 estimates of the labor force and agricultural output linked to these two series. The points clarified concern the effects of the westward movement and foreign trade upon the course of per capita income. David uses Parker's crop estimates and a heretofore unpublished set of estimates prepared by Easterlin showing 1840 agricultural output by regions in uniform national prices. With them, he answers Parker-Whartenby and Taylor concerning possible adverse effects of the westward movement. Similarly, his use of North's data on foreign trade and the terms of trade permits him to strongly argue that North and Taylor have considerably overstated the case concerning the effect of trade on income movements.

David's conjectures are used to argue against two hypotheses concerning economic development in the first half-century of nationhood. Because he is attacking conflicting hypotheses, David's discussions of possible biases in the results are somewhat misleading. What is a bias against one conclusion can be a bias in favor of the other. However, a discussant is placed in the comfortable position of just raising questions without concern about their effects upon the outcome of the debate. The first attack is on the conclusion, accepted by Martin, North, and Taylor, that per capita income was roughly the same in 1840 as it was in 1800. David marshals strong evidence to argue that such an occurrence was unlikely. The reasons are those first suggested by Kuznets—a shift within the labor force to the higher productivity (per worker) nonagricultural sector and a rise in the ratio of labor force to population. The relative importance of these differ in David's estimates, based as they are upon a different set of labor force estimates. Thus even if labor productivity within each sector had remained constant, per capita income would have risen; and David uses evidence based on wage and price data to support the claim of intrasectoral labor productivity increases.

The second argument is with those such as Rostow and Goldsmith who assert that there occurred a sharp acceleration in the rate of growth of per capita income at some time in the antebellum period. David claims that there was no such acceleration between 1790 and 1860, and suggests that no upsurge occurred prior to that period. On this point, although my intuition would be to believe no sharp break did occur, less confidence can be placed in inference based upon the income estimates. If we eliminate the decade of the 1790's, acceleration in the growth rate did occur after 1820, with the long swing

pansion of 1820-34 being one of the most rapid periods of growth in U.S. history. (It should be noted that this is the period stressed by Fogel in his critique of Rostow.) However, David points to a high estimated growth rate between 1790 and 1800, making the 1800-20 interval a lull between two periods of rapid growth. Given the heavy reliance upon the structural change in the distribution of the labor force for the 1790's estimate, one would prefer stronger supporting evidence. Such evidence is undoubtedly unavailable, but the magnitude of the shift into nonagricultural employment seems high in light of the not abnormal urbanization of that decade. David's estimates imply a growth rate of nonagricultural output of about 10 percent per annum, which must be among the highest rates recorded and exceeds that in any pre-1860 decade. And given the agricultural labor force shares shown in the time series and cross-sections collected by Kuznets, if the 1790 estimates are correct, it is difficult to visualize any such marked growth rates in the pre-1790 period.

There are several surprising implications in the conjectural estimates which, if true, will force revisions in the way we look at the pattern of U.S. development. Since the Towne-Rasmussen estimates for 1800-40 rest almost entirely upon the assumption of constant per capita domestic consumption, the export figures having only minor effect, there is no trend in per capita agricultural output before 1840. Gallman's data for 1840-1900 show only a small increase in per capita agricultural output, the one irreversible rise occurring in the decade of the 1870's. Thus, abstracting from fluctuations, the data suggest about a zero income elasticity of agricultural output per capita for a period of 110 years. The Lebergott series, which in a sense is the key to David's findings, presents a much slower growth in the agricultural labor force from 1800 to 1840 than do the series used by previous writers; 2.4 percent per annum versus, e.g., 3 percent in the series used by Towne-Rasmussen. It is the relative decline in the share of the labor force in agriculture with the assumption of constant per capita consumption which yields David's estimates of increased output per worker in agriculture and nonagriculture. (Use of earlier agricultural labor force estimates in conjunction with Lebergott's participation rates would show almost no measured growth in per capita output. Given our proclivity to justify what data appear to reveal, perhaps some tolerance should be granted those who used the earlier figures.) While the Lebergott series is clearly superior to earlier variants, it does imply that a substantial part of the nineteenth-century shift out of agriculture was concentrated in the period prior to 1850, the years from 1850 to 1880 showing a small decline. In David's estimates there results an implied rate of growth of agricultural output per worker from 1810 to 1840 in excess of the rate of increase for the rest of the century! The 1820-30 increase in agricultural productivity (a decade in which Easterlin's data show a small contribution of the westward movement) is comparable with the peak decades of the later part of the century. Given the emphasis on agricultural mechanization, these are surely surprising inferences.

I have difficulty in interpreting David's description of the procedure used to extrapolate income estimates. David computes the 1840 relative sectoral productivity per worker and, for computational purposes, holds this ratio constant over time. To reduce the measured effect of intersectoral shifts he uses

the variant with the higher agricultural relative. However, he then states that his other decade estimates are based upon the less inclusive concept of GDP. While one can understand his desire to provide a unique measure of the effect of the intersectoral shift, as well as his attempt to minimize that effect, his procedure is misleading. Based on his formula, all that matters are changes in the variables, the 1840 absolute level being irrelevant except to fix relative sectoral productivities. Once this ratio is fixed, consistent totals of sectoral outputs can be derived only by using that variant used to set the ratio. The use of another variant builds in an implicit trend in relative sectoral productivities which is not obvious, and here implies a relative increase in product per worker in agriculture after 1800. The two concepts discussed by David did grow at different rates, but that does not justify the procedure, and the concept David used should yield higher growth rates than would the more inclusive measure. While estimates permitting changes in relative output per worker in the different sectors may be better than those assuming a constant ratio, it would seem better to have such assumptions explicitly introduced.

Finally, how comfortable can the historian be with David's conjectural estimates? While the extrapolation of certain ratios may permit the setting of bounds and subsidiary data indicate the plausibility of such bounds, the procedure biases the conclusion for those who would argue that there were discontinuities in the process of growth. David's results rest heavily on the assumptions of near constant per capita agricultural output between 1790 and 1840 and a constant ratio of output per worker in agriculture relative to non-agriculture (at the 1840 level). For those who argue that these ratios fluctuated or changed, no matter which conclusion they wish to defend, David has put aside much of the controversy. When David notes the success of his prediction formula for the 1840 to 1860 period, he is basically showing the stability of one of the ratios, relative sectoral productivities, but since he is using actual output estimated by Gallman, the Towne-Rasmussen constant per capita consumption assumption is not tested. The latter would seem as crucial an assumption as the one tested by David.

All of these comments do not do justice to the meticulous care with which David has prepared his estimates, justified them, and analyzed the literature. His conjectures will form a necessary starting point for the further analysis of these early years of economic development in the United States.

There are two points I should like to raise about Professor Scoville's interesting paper that are relevant to any study of one particular industrial sector. First, what was happening to the size of the market for the industry? Perhaps there is no data available to determine the overall growth of the French economy, but it would be interesting to determine the effects of the loss of foreign markets due to tariffs. If Coleman's contention about the sheltered nature of the English industry is correct, it may be that the relative difficulties of the French industry are overstated. Second, Scoville's discussion of the disadvantages of government interference and the presumed beneficial effects of liberalization may be open to some question. Such liberalization as did occur was still within the confines of a protective tariff and, perhaps more importantly, heavy and apparently increasing control over the raw material, rags, which formed a large proportion of total costs. The export of rags was prevented, ei-

ther by explicit provision or prohibitive duty, and the control over supply led to restrictions upon the location of raw material stores and even the location of the paper establishments themselves. Perhaps the implication is that appropriate government policies of control over inputs can lead to rapid growth of final output of an industry, since it would seem that people dealing in rags did not benefit from any liberalization.

Ross M. ROBERTSON: These papers have been particularly rewarding, largely because for the first time in several years we have listened to essays in economic history rather than about economic history. These papers suggest that there are not really two kinds of economic history, as frequently alleged, but rather just one, based on solid scholarship and aimed at finding out the truth about the past.

As Professor Scoville suggests, anyone familiar with Professor Hamilton's work knows that he is not afraid of marshaling data. (The British, incidentally, are still arguing about conclusions he reached in his seminal work published more than thirty years ago.) Scoville, though cautioning us about the validity of his own data, nevertheless gives us a good notion of quantitative change in the French paper industry during the eighteenth century.

Yet each of these first two papers deals with ideas rather than quantities. Hamilton devotes a large portion of his paper to two points. The first, that John Law was not a compulsive or professional gambler, is a result that we might have expected. The second, that Law was a vindictive man, continually exhibiting hostilities that would bring him to ultimate ruin, is a little harder to believe. Yet assuming that Hamilton has chosen adverbs with his usual care, we can only conclude that Law's disintegration began inside the man rather than outside.

I find Scoville's paper on the whole persuasive. I certainly have no quarrel with his quantitative inferences, nor for that matter with his description of French governmental regulation in the eighteenth century. I was not altogether convinced of the direct correlation between the paper industry's expansion and the relaxation of direct governmental controls that he alleges. Perhaps Scoville's more extended publications on this subject will elucidate this relationship.

I must confess to a special interest in Professor David's paper on real product growth in this country before 1840. Like many another middle-aged historian in this room, I recall how puzzled we were when Robert F. Martin first published his income and product estimates in the late 1930's. Even if we allowed for errors and for the fact that some of the census years were depressed and others prosperous, it appeared that per capita income of Americans declined steadily between 1799 and 1829. Converting Martin's series of private production income per capita to an index (1799 = 100), we have:

	1799	1809	1819	1829	1839
Current dollar series.....	100	99	71	58	74
Cost-of-living series......	100	96	79	77	92

From these data it could only be concluded that growth of output was particularly slow during the 1810's and 1820's.

Professor David reports the concern that Kuznets and, later, Parker and Whartenby expressed about the Martin estimates. In formulating his theory of a "take-off" of the American economy during the 1840's, Professor W. W. Rostow based his propositions more on twentieth-century experience than on nineteenth-century evidence. Actually, it was Professor Robert E. Gallman's calculations, published in the same volume with the Parker-Whartenby results, that cast the final shadow on Martin's work, for Gallman's estimates indicated astonishingly high decennial increases in the value of commodity output for the period 1839-59. Indeed, a rate of growth for this period of more than 50 percent in the value of commodity output made it almost impossible to believe that the orthodox view of growth rates between 1789 and 1839 was correct.

Basing my calculations on changes in the balance sheets of selected financial institutions, I wrote some years ago that income per capita may have increased by as much as two-thirds over the fifty-year period 1789-1839.[1] It was possible then to hypothesize that the rate of growth of the American economy did not jump suddenly in the 1840's, but instead accelerated only slightly between 1790 and 1860. Such a hypothesis, if substantiated, meant that there would have to be a revision of our ideas about the way growth took place in the United States.

It seems to me that Professor David's evidence is convincing. He concludes that by 1840 real gross domestic product, per capita, was probably 60 percent higher than it had been at the beginning of the century. With the tendency of the American economy toward pronounced economic fluctuation in the nineteenth century, we can readily accept the observation that the rate of change was not steady. On the other hand, the older historians exaggerated the swings in output experienced in the first half of the nineteenth century. When we reflect that interchangeable-parts and continuous-process manufacture were pre-1800 innovations in this country, that the factory system was introduced in the textile industries before 1815, and that there was tremendous investment in canal building in the 1810's and 1820's, we should not after all be astonished at a more rapid rate of growth than previously suspected. When we reflect further that the commercial banking system was ideally devised to promote growth and that central-bank experimentation, modified by the vulgar profit motive, did little to inhibit free-wheeling early entrepreneurs, we are even more readily persuaded by David's pathbreaking paper.

These findings will of course be refined and modified. I am disposed to think, however, that further inquiry into colonial rates of growth may lead to the conclusion that the American economy grew at approximately the same rate for three hundred years. If it did, some fashionable growth models, even now of dubious value, will have to be changed.

[1] See Ross M. Robertson, *History of the American Economy* (Harcourt, Brace and World, 1964), p. 239.

ECONOMIC DEVELOPMENT

SOME LESSONS OF HISTORY FOR DEVELOPING NATIONS*

By RONDO CAMERON

University of Wisconsin and
Instituto de Economia, Universidad de Chile

In recent years economists have recognized that historical experience can be a valuable guide for both theory and policy [2]. Nevertheless, many economists and policy-makers, especially in developing nations, persist in believing that the historical experience of modern industrial nations is irrelevant to the problems faced by developing nations [23] [30]. The purpose of this paper is to argue the contrary proposition. This is not to say that contemporary developing nations will inevitably follow the same pattern of development as the highly industrialized nations, much less that the policies of the nineteenth century are appropriate in the vastly different circumstances of the second half of the twentieth century. Historians would be the last to deny that there are alternative routes to economic development, or that different conditions require different policies. Even Marxist economists no longer believe Marx's dictum that underdeveloped nations can see in the history of the developed ones the mirror image of their own future. This does not signify, however, that nothing at all can be learned from history. There are, in fact, "lessons of history" as abstract from time and place and almost as ineluctable in their import as a multiplication table. For policy-makers in developing nations to ignore them—as they do as often as not—is to condemn themselves to failure and their nations to perpetual poverty.

The following brief discussion is limited to four areas of major importance for contemporary policy-makers. Limitations of space force me to omit other areas scarcely less important, such as the role of agriculture in industrialization, and to condense drastically the available evidence on the topics mentioned.

The Role of Capital

Economists from Marx and Engels to Harrod and Domar, as well as earlier and more recent ones, have stressed the preeminent importance of the role of capital in economic growth. After World War II the almost unanimous and sometimes almost the only prescription of economic advisers from developed nations for those beginning to industrialize was massive injections of capital [20] [23] [24]. Rostow's celebrated take-off hypothesis, which emphasized a rapid rise in the rate of net investment from 5 percent or less of national income to 10 percent or more in order to achieve sustained growth, appeared to give empirical support to the theoretical propositions. In a number of countries the prescription of the capital intensive economists was carried out through a combination of capital imports and forced savings by means of government finance. Although a few countries have achieved relatively high rates of growth, the results as a whole have been far from satisfactory; nor is there any clear correlation between investment ratios and rates of growth [29].

Criticism of the take-off hypothesis based on recent historical research has virtually eliminated the key role assigned to the rapid rise in the investment ratio. This research indicates that almost every developed country of today entered a phase of sustained growth with investment ratios substantially below the magic figure of 10 percent; and that the rise in that ratio followed rather than preceded the adoption of new technologies. For Great Britain the work of Phyllis Deane and her collaborators shows that for most of the eighteenth century the rate of net capital formation measured as a percentage of national income did not rise above 5 percent, reaching possibly 7 or 8 percent only in the last decade of the century; not until the railway boom of the 1840's—that is, well after the industrial revolution proper—did the ratio rise to 10 percent [8] [9] [10]. Moreover, according to Miss Deane, during the full flush of the industrial revolution at the end of the eighteenth century "the annual flow of new capital into the leading commercial and industrial sectors . . . was not more than . . . perhaps 1 percent of national income" [8, p. 366]. Agriculture, transportation, and urban construction each absorbed more capital than mining and manufacturing together.

Similar situations appear to have obtained in other countries during the early stages of industrialization, although for most of them the statistical evidence is sparse and unreliable. Preliminary results obtained by Jean Marczewski's team suggest that in France net capital formation as a percentage of net domestic product did not rise above 3 percent per annum, on the average, until the 1840's, and did not rise

above 10 percent until the railway and urban construction boom associated with the Second Empire [19, p. 121]. The 3 percent figure appears to be implausibly low, and is probably due to imperfections in the basic data; but Marczewski himself admits that his estimates for the second half of the century are probably too high.

The earliest figures for Germany pertain to the 1850's, when German industrialization was already in full swing, and indicate a ratio of approximately 7.5 percent [15, pp. 114, 115]. Professor Hoffmann, the leading authority, regards 5 percent as a "realistic figure" for the period 1831-55. In Sweden, the country in which the process of industrialization is best documented statistically, the ratio apparently did not exceed 6 or 7 percent until the twentieth century [5] [16]. For Japan, finally, the best estimates seem to indicate that the ratio of net investment to national income remained roughly constant at 7-8 percent from the 1880's until the early years of the twentieth century; not until after World War I did the ratio rise appreciably [26].

What is the significance of these data for theories of economic growth which assign such great importance to the factor capital? In the first place, it seems clear that the increase in the investment ratio is a consequence rather than a cause of economic growth; rich nations, like rich individuals, tend to save and invest a larger proportion of their incomes. Second, capital is not, as we used to be taught, a homogeneous factor; this seems obvious to economic historians, but it is only recently, as a result of developments on other fronts, that we have become aware of the importance of the heterogeneity of capital. In particular, capital as conventionally measured does not take adequate account of investment in human skill and intelligence. Taking this into account, it seems likely that today's industrial nations began modern economic growth with substantially larger stocks of capital than would appear from the statistical evidence. This is a point to which I shall return. Important as these considerations are, however, they do not explain how economic growth began, nor the precise role of capital in initiating growth. From historical evidence it appears that the key development in getting the growth process under way was a more intensive and efficient use of both the existing stock of capital and of normal increments to that stock.

It is easy to demonstrate theoretically that an economy can experience an increase in total product per worker even with zero net investment by means of the replacement of worn-out capital equipment with new equipment having a higher productivity as a result of technological improvements. Similar results can be obtained as a result of organizational improvements or financial and administrative innovations. Both processes appear to have been at work during the English in-

dustrial revolution and similar periods of rapid change elsewhere. So simple an invention as the spinning jenny—the first of a series of innovations that revolutionized the cotton industry—permitted one man to do the work of sixteen. The adoption of the factory system increased productivity through a more intensive use of capital equipment and by permitting more continuous and efficient supervision of labor. In connection with improvements in transport (turnpikes, canals, railways), it also permitted a reduction in the proportion of working capital required under the domestic system of production. The adoption of steam power and artificial illumination freed early factory industries from reliance on the raw forces of nature and allowed more continuous operation regardless of the season.

Financial innovations associated with the rise of banking, including the use of banknotes and demand deposits as stores of value and means of payment, economized on working capital and permitted a more efficient allocation of savings to investment [7]. Traditional economies typically "lock up" a large proportion of their meager stocks of capital in such unproductive forms as treasure hoards (including hoards of commodity money) and excessive inventories of foodstuffs and other commodities. The introduction of financial intermediaries allowed this "primitive accumulation" to be used more productively by providing holders of idle capital with a wider variety and more attractive range of financial assets, as well as by giving entrepreneurs more convenient access to loanable funds.

The same considerations of intensity and efficiency apply to the use of foreign as to domestic capital. By and large, foreign capital in underdeveloped nations which is not accompanied by the technical and organizational skills to put it to effective use, as is the case with the large majority of loans to governments, is unlikely to make much of a contribution to economic development. The results of the relatively small investment of French capital in Belgium and Germany in the mid-nineteenth century stand in marked contrast to the much larger investments in Spain, Italy, and Russia [6].

What are the lessons in this for contemporary developing nations? Under the influence of outmoded economic theories, policy-makers in many developing nations have developed a naïve faith that a rise in the investment ratio is a necessary if not a sufficient condition for economic development. It should be made clear to them, therefore, that such a rise is neither sufficient nor necessary. While in most cases an increase in the rate of real capital formation would be desirable, it does not follow that development cannot take place by means of a more efficient allocation and use of existing stocks of capital and normal increments thereto. Almost every underdeveloped country has its

examples of idle, underutilized, or misallocated capital: extravagant public buildings and similar monuments, such as athletic stadia; unprofitable "prestige" airlines; large modern factories which remain idle or work at less than full capacity because of poor planning, bureaucratic interference, lack of markets, shortages of raw materials, etc. Most developing nations hold excessive inventories (relative to the flow of total product) of both producers' and consumers' goods as a result of slow and uncertain deliveries from both domestic and foreign suppliers due to poor transport facilities, exchange controls, and administrative delays. Even so, the periodic temporary exhaustion of raw materials and spare parts results in the enforced idleness of much capital equipment. In short, there are still many economies to be achieved. The emphasis in development planning should shift, I suggest, from attempts to raise the investment ratio at whatever cost, to attempts to ensure a more rational, efficient, and intensive use of existing capital.

The Role of Labor

Whereas historical experience with respect to the role of capital is encouraging for today's developing nations, that with respect to the role of labor is much less so. Greater output through a more intensive and efficient utilization of capital carries with it an important corollary; namely, a more efficient and intensive utilization of labor. The length of the working day during the English industrial revolution—twelve or fourteen hours for women and children as well as for adult males—was one of its characteristics most deplored by contemporary reformers as well as by latter-day historians. Nor were such hours peculiar to England; all industrializing nations have experienced them. As recently as the early years of this century the average workday in manufacturing in the United States was just under ten hours, having been reduced from twelve to fourteen in the mid-nineteenth century [17]. In agriculture the average workday was, if anything, even longer.

Such facts are not very agreeable, but they cannot be wished away. I find it surprising that so little attention has been paid to them in discussions of development policies and programs. On the contrary, faced with the demands of organized and/or politically conscious workers, the development policies of most noncommunist nations call for a reduction in the length of the workday or workweek (along with minimum wages and large social insurance benefits on the model of their more affluent neighbors) at the same time that they are striving for a larger output per person. The situation is further aggravated in most developing nations by the existence of substantial pools of unemployed or underemployed workers in both urban and rural areas. This is, to say the least, a very horny dilemma.

I want to make it clear that I am not advocating a return to nine-

teenth-century labor standards. Indeed, there is much evidence, historical and other, that the excessive length of the workday characteristic of early industrialism did not necessarily represent the most efficient use of labor. On the other hand, there are many ways to achieve a more productive use of labor without resorting to crude exploitation. One of the definitional characteristics of an underdeveloped country is the existence of a large, underemployed, low-productivity labor force in agriculture. Evidence from the United States, Japan, and several other countries of the large role of agricultural labor in the formation of agricultural capital is therefore of some relevance [21]. Achieving a more productive use of labor in this fashion will probably require some changes in systems of land tenure—but that is a subject too vast and complicated to be entered into in so brief a disquisition. It is also likely that to secure a more intensive and efficient application of labor in agriculture and elsewhere will require a prolonged process of education, for it is by now obvious that the "propensity to truck and barter" and other aspects of economizing behavior which Adam Smith and his disciples took for granted is not instinctive but culturally conditioned. It is, in short, "learned behavior." The inhabitants of the underdeveloped world have learned to desire and demand the outputs of modern economic systems, but they have not yet learned, or have done so only imperfectly, the much more complicated relationships between inputs, efficiency, and outputs. This dictum applies to entrepreneurs and policy-makers as well as to workers, to private and public systems of incentives and rewards as well as to politically-motivated work stoppages.

Policy planners in developing nations will argue that the higher productivity of labor made possible by modern technology will allow them to enjoy rising per capita incomes without resorting to the "exploitation" of workers that characterized the earlier history of developed nations. That may be so, but modern technology also requires a skilled, literate, and responsible labor force—not to mention intelligent and resourceful entrepreneurs—to utilize it effectively.[1] Until such time as the developing nations can educate and train their labor forces and entrepreneurs to much higher standards than exist today—and for most of them that time is not imminent—they face the bleak prospect of continued low productivity and incomes.

The Role of Education

In recent years the proliferation of studies in the economics of education has convinced most economists of the importance of human cap-

---

[1] It is frequently observed that the technologies developed in industrially advanced nations are not always applicable or advisable for developing nations without substantial modifications. In most cases the modifications involve a smaller scale of enterprise and more intensive use of labor.

ital for the economic growth of advanced industrial nations [11] [12] [13] [27] [28]. As yet, however, there have been few if any studies of the relationship between education and the beginnings of development. Did today's wealthy countries begin the process of modernization with relatively advanced educational systems? Or on the contrary, were they able to create elaborate educational systems with the increments of wealth produced in the process of growth? While there is little doubt that, once growth has begun, the process is one of mutual interaction—knowledge begets wealth, poverty begets ignorance, and vice versa—one would like to know the nature of the relationship at some "time-zero." Anderson and Bowman state, in their analysis of needed research on the role of education in development: "One of the main questions calling for investigation is that of the interplay between the spread of literacy and elementary education in the early stages of the agricultural and industrial revolutions in Western countries and Japan" [3, p. 173]. While I have not made the detailed investigations called for, I would like to present some observations based on general studies of the economic history of the developed nations. (In this as in all other sections of the paper the documentation suffers from the fact that it is being written in one of the developing nations, where library resources are less than overwhelming.)

Britain at the end of the eighteenth century probably had the highest literacy rates in the world at that time. It is estimated that in the 1840's 60 percent of adult males and 40 percent of adult females could read and write; and there is some evidence that literacy was even higher at the end of the eighteenth century than in the 1840's [14]. The special case of Scotland is of exceptional interest. The superiority of the Scottish educational system over that of the English, at every level from primary schools to universities, dated from the Reformation. In the course of the eighteenth century every parish came to have at least one school. Some of the most notable technical and organizational innovations of the industrial revolution originated in Scotland, and in the century from 1750 to 1850 the rate of economic growth in Scotland far outstripped that of England, converting Scotland from a primitive tribal economy into one of the most heavily industrialized of the times [7, Chap. III].

Germany at the end of the eighteenth century was a poor, backward area, in spite of the success of Frederick the Great in building up the military power of Prussia. A more fruitful if less spectacular accomplishment of that ruler was his creation of the nucleus of a broadly-based system of elementary education. In the campaign of reforms stimulated by the humiliating defeat of Prussia in the Napoleonic Wars that system was extended, and in subsequent decades technical

institutions were established, the universities reorganized and revitalized, and new universities were created. By the 1870's German universities were recognized as the world's best, especially in science and engineering, whereas half a century previously they had been either tradition-bound or imitative [6, pp. 52-53]. The important point, however, is that the restructuring and extension of the educational system took place prior to the advent of rapid industrialization.

Similar phenomena occurred in Switzerland, the Netherlands, and the Scandinavian countries, and help to account for their rapid industrialization at the end of the nineteenth century. Limitations of space prevent me from discussing the role of education in U.S. economic development, which in any case should be more familiar to this audience. For the same reason I refrain from commenting on the particular strengths and weaknesses of the French educational system, which I have dealt with elsewhere [6, Chap. III]. One final case should not be overlooked, however. According to R. P. Dore, Japan started "its career of forced-pace modernization with a widespread and well developed tradition of formal institutionalized education" [25, p. 105]. Professor Rosovsky declares that on the eve of the Meiji Restoration of 1868 literacy rates were higher "than in any European country at a comparable stage of development, with the exception of Holland and Prussia " [ibid.]. Although I would like to insert more exceptions, such as Switzerland and the Scandinavian countries, at least on a provisional basis until I see more evidence, the evidence already available is sufficient to justify Rosovsky's assertion that "Japan's stock of education—and human capital—was unusual by international standards" [ibid.]. Moreover, the policies of the Meiji government worked to increase those stocks. According to Lockwood, "no other enterprise of the State paid handsomer dividends to the nation" than its investment in education [18, p. 512].

This lesson for developing nations is unusually clear, I believe. A modern, widespread, and variegated educational system, far from being a luxury which only wealthy nations can afford, is more of a necessary cost which developing nations must incur if their plans and aspirations are to be achieved.

The Role of the State

We arrive, finally, at the most controversial area. What is the optimal role of the state in contemporary developing nations? How does it differ, on the one hand, from the role of the state in those countries that began to industrialize in the nineteenth century on a predominantly private enterprise basis, and, on the other, from those countries—principally the Soviet Union—that have succeeded in building a more

or less modern industrial structure on the basis of comprehensive state control and planning? The subject is certainly highly emotive as well as complicated; it may be foolhardy even to mention it in such a brief and necessarily inconclusive fashion. I wish to emphasize, therefore, that my purpose is not to define the optimal role of the state on the basis of historical experience but merely to indicate how that experience may be relevant even under the vastly different circumstances of present-day developing nations.

The "choice" between comprehensive state control and a less closely regulated economy is not in any meaningful sense of the word a choice at all. For any given country the outcome will be determined by forces that are not amenable to rational decision making. For that reason I restrict my remarks to cases in which rational choice with respect to the nature and extent of state participation in the economy is still possible. I do not mean to imply that Soviet-style economies cannot also utilize rationality in decision making, or benefit from historical experience, nor that other economies cannot learn from their experiences—for example, the inevitability of coercion and regimentation which such systems involve—merely that the subject is sufficiently complicated as it is.

In order to keep ideological interferences to a minimum, let us begin by clearing the ground of old myths and half-truths. It is not true, in the first place, that the classical economists advocated a complete "hands off" policy for the state [22]. Second, it is even farther from the truth to suppose that nineteenth century governments adhered systematically to the doctrines of the economists in that or any other matter. These points are so well established that it may appear gratuitous to repeat them to this audience; I include them for the benefit of the wider audience for whom this paper is ultimately intended.

The exact nature and extent of government participation in the economy varied considerably from country to country and from one period to another within any given country. Britain came the closest to achieving the program of the economists (though one should remember that economists, then as now, were rarely unanimous on any specific policy proposal); but Britain was also the home of Benthamite liberalism, the forerunner of one important collectivist school of thought. At the same time that the government was engaged in dismantling the old clumsy apparatus of restrictive legislation and institutions, it began a whole new series of legislative interferences with individual activity in the name of "the general welfare" or for the protection of the weak and disadvantaged [4].

Outside of Britain governments generally took a far more active role

in the economy, from the decentralized and competitive promotional activities of state and local governments in the United States during the Jacksonian era to the highly centralized and bureaucratized activity of the Russian state under Count Witte in the 1890's. The results, as one might suppose a priori, were highly varied; some enterprises and activities succeeded admirably; others were miserable failures. Thus one cannot generalize completely about the efficacy of government participation from historical experience.

Government participation in the process of production (I leave aside for the moment welfare activities) may take one of two broad forms: (1) it can encourage and stimulate private enterprise, strengthen the operation of free markets, and supplement and correct them when they are too weak or inadequate to produce the desired results; or (2) it can seek to control or eliminate private enterprise and to restrain or replace free markets by government edicts. Generally speaking, government policies in those countries that succeeded in industrializing, including Japan and Russia under Witte, were of the former type, whereas those in countries that notably failed to industrialize, such as Spain and the Ottoman Empire (and Russia for most of the nineteenth century), were of the latter type.

It is my impression that the policies of many developing nations resemble primarily those of the second category mentioned above. In part this is understandable, even if the policies themselves are not wise. Frequently, especially in ex-colonial areas but also in many countries of Latin America, private enterprise is identified with foreign investment, which, whatever its contributions to domestic well-being may or may not have been, is regarded with suspicion. In other cases private enterprise of the traditional type has too often been unenterprising, the ally or offspring of traditional ruling classes and/or corrupt governmental patronage. Whatever the sources of resentment against private enterprise and however earnest and idealistic the desires of government officials of developing nations to better the lot of their fellow citizens under the tutelage of the state, they should consider carefully and dispassionately the possibilities and limitations of direct versus indirect action on the economy. The forces of the market are powerful, even within backward, underdeveloped economies. Harnessed in the interests of society by means of appropriate fiscal and monetary policies and similar instruments, they can contribute importantly to economic growth and development. If, however, the state attempts to overrule or outlaw them, they can produce chaos and stagnation. All too often policy-makers attempt to operate directly on the target variables, where their action is not only inefficacious but down-

right harmful to the cause of economic development; for example, by attempting to curb inflation by means of price controls, while simultaneously expanding the stock of money.

Many of the alleged deficiencies of private enterprise in underdeveloped countries are the result of government activities—frequently contradictory and/or mismanaged—that are justified by references to those very deficiencies. For example, manufacturing industry is encouraged because of its apparently higher productivity than agricultural employment; but the differential in productivity may result at least in part from tariff protection and other favors granted to manufacturing, while agriculture stagnates with inadequate transport, marketing, and credit facilities, and positive disincentives in the form of price controls. On the other hand, industrial entrepreneurs fail to exhibit dynamism because they rest secure in the knowledge that exorbitant tariffs will protect their limited domestic markets but are unable even to try to compete in foreign markets because of tariffs on their raw materials, overvalued currencies, and other disincentives to export.

It is frequently argued that the shortage of capital and entrepreneurs in contemporary underdeveloped economies is the reason why the state must play a larger direct role than was true in the nineteenth century. Such arguments are unsound historically. As I have already indicated, many if not most of today's developing nations have higher rates of capital formation than was the case with the developed nations in their early stages of industrialization. As for the shortage of entrepreneurs, that argument can only be tested after the fact. There was no reason to expect that Scotland in the mid-eighteenth century or that Japan in the 1860's would soon produce some of the world's most resourceful entrepreneurs.

Arguments for a larger direct entrepreneurial role for the state also overlook the fact that decisions and actions must still be taken by men. The kinds of men to be found in the service of the state—and, more importantly, the kinds of incentives that impel them—are not necessarily those most conducive to economic growth. By their very nature, political organizations attract many individuals who have no qualifications for getting a job done except their devotion to the party organization and (all too often) their desire to wield power and dispense patronage. Insofar as there is an actual shortage of trained, skilled, and dedicated manpower, it will be found to be at least as acute in government as in the private sector—and with a far greater potential for harmful effects. Even the ablest of individuals will be rendered ineffective if burdened with too many tasks that cannot be delegated for want of capable subordinates. In the circumstances of most developing nations, Parkinson's Law is no laughing matter; it is

a frightful reality. Many examples, which would be amusing if they were not so tragic, could be cited.

There is little doubt that government officials today are, in general, much more "growth-minded" than their counterparts in the past, but that is not necessarily a valid reason for substituting their energies and decisions for those of private entrepreneurs. It is a common observation that state-sponsored development projects tend toward the grandiose and monumental, frequently of dubious economic value, while small-scale industry and agriculture are often neglected. The most successful historical cases of industrialization, however, show the predominance of small-scale enterprise in the earliest stages of industrialization.[2] If small enterprises are successful, it is because they have met the test of the market, and they will inevitably grow. On the other hand, if they fail, the losses to society are much less than in the case of large government-sponsored enterprises, which for political reasons may not be allowed to appear to have failed.

Let me reemphasize that I am not advocating a laissez faire role for the state. Government has, indeed, many important functions in modern economies, both developed and developing—but the wholesaling of onions and potatoes is not one of them. In addition to maintaining an "exact system of justice"—no small task in itself, and one which is very badly performed in many underdeveloped countries—there are huge tasks ahead in almost all developing nations in providing more adequately for the health and educational needs of their populations, as well as maintaining other essential services which government can do more effectively than private enterprise. These tasks are quite sufficient to employ fully both the fiscal and human resources which most governments can command; until they can operate their railways and postal systems efficiently, they should not consider a petrochemical industry, much less the wholesale distribution of foodstuffs.

In closing I would like to append a yet more personal comment. I am aware that the views expressed in this paper, and especially those of the concluding section, are lacking in scientific rigor. Whether for good or ill, policy making is still more of an art than a science, and discussions of policy necessarily reflect the personal biases of the discussants. Let me make it quite clear, therefore, that my own biases are not based on any doctrinaire ideological grounds whatsoever. On the contrary, as an "old socialist" (or, at any rate, a middle-aged former socialist), whatever ideological proclivities I retain run in the opposite direction. My "conversion," if that is the proper term, is the result of twenty years of intensive study of the historical record and a somewhat shorter but no less intensive exposure to the vagaries of policy

[2] The considerations of fn. 1 apply to this point also.

making in developing nations. History, as I have suggested, is a great teacher—personal history not least of all.

REFERENCES

1. M. Abramovitz, "Resource and Output Trends in the United States since 1870," *A.E.R.*, May, 1956, pp. 5-14.
2. *A.E.R.* "Economic History: Its Contribution to Economic Education, Research and Policy," May, 1965, pp. 86-112.
3. M. J. Bowman, and C. A. Anderson, "The Role of Education in Development" in R. E. Asher, *et al., Development of the Emerging Countries: An Agenda for Research* (Brookings Inst., 1962).
4. J. B. Brebner, "Laissez-faire and State Intervention in 19th Century Britain," *J. of Econ. Hist.*, 1948, pp. 59-73.
5. A. K. Cairncross, "Capital Formation in the Take-off," in *The Economics of Take-off into Sustained Growth*, W. W. Rostow, ed. (London, 1963), pp. 240-60.
6. Rondo Cameron, *France and the Economic Development of Europe, 1800-1914* (Princeton, 1961; rev. ed., Chicago, 1966).
7. ———, *et al., Banking in the Early Stages of Industrialization* (New York, 1967).
8. P. Deane, "Capital Formation in Britain before the Railway Age," *Economic Development and Cultural Change*, No. 3, 1961.
9. P. Deane and W. A. Cole, *British Economic Growth 1688-1959, Trends and Structure* (Cambridge, 1962).
10. P. Deane, *The First Industrial Revolution* (Cambridge, 1965).
11. E. F. Denison, *The Sources of Economic Growth in the United States and the Alternatives before Us* (C.E.D., 1962).
12. S. Fabricant, *Trends in Income and National Product in the U.S.* (N.B.E.R., Occasional Paper No. 63, 1959).
13. F. Harbison and C. Myers, *Education, Manpower and Economic Growth* (New York, 1964).
14. J. F. C. Harrison, Letter to the author, Madison, Wis., Aug. 17, 1966.
15. W. Hoffman, "The Take-off in Germany," in W. W. Rostow, ed., *op. cit.*, pp. 95-118.
16. S. Kuznets, "Notes on the Take-off," in W. W. Rostow, ed., *op. cit.*, pp. 22-43.
17. Stanley Lebergott, *Manpower in Economic Growth: The American Record since 1800* (New York, 1964).
18. W. W. Lockwood, *The Economic Development of Japan* (Princeton, 1954).
19. J. Marczewski, "The Take-off Hypothesis and French Experience," in W. W. Rostow, ed., *op. cit.*, pp. 119-38.
20. R. Nurkse, *Capital Formation in Underdeveloped Countries* (New York, 1962).
21. M. L. Primack, "Farm Capital Formation as a Use of Farm Labor in the United States, 1850-1910," *J. of Econ. Hist.*, 1966, pp. 348-62.
22. L. Robbins, *The Theory of Economic Policy in Classical Political Economy* (London, 1952).
23. P. N. Rosenstein-Rodan, "Problems of Industrialization in Eastern and Southeastern Europe," *Econ. J.*, 1943, pp. 202-11.
24. ———, "International Aid for Underdeveloped Countries," *Rev. of Econ. and Statis.*, May, 1961, pp. 108-36.
25. H. Rosovsky, "Japan's Transition to Modern Economic Growth, 1868-1885," in H. Rosovsky, ed., *Industrialization in Two Systems: Essays in Honor of Alexander Gerschenkron* (New York, 1966).
26. ———, *Capital Formation in Japan, 1868-1940* (New York, 1961).
27. T. W. Schultz, "Investment in Human Capital," *A.E.R.*, 1961, pp. 1-16.
28. ———, "Education and Economic Growth," in H. G. Richey, *Social Forces Influencing American Education* (Chicago, 1966).
29. United Nations, *Yearbook of National Accounts Statistics* (1965).
30. C. Veliz, ed., *Obstacles to Change in Latin America* (London, 1965).

SOURCES OF POSTWAR GROWTH IN NINE WESTERN COUNTRIES

By EDWARD F. DENISON
Brookings Institution

This session, according to its organizer, was designed to see what individuals studying growth by different approaches have learned. My approach is to try to isolate the sources of observed growth of real national income between selected dates and quantify their contributions. The contribution of each source is expressed as the number of percentage points it contributed to the growth rate.

I used this approach a few years ago to analyze United States growth from 1909 to 1957.[1] My paper today concerns an application of similar techniques to nine countries in the 1950-62 period and 1950-55 and 1955-62 subperiods. J. P. Poullier assisted me throughout this study. The countries covered and their growth rates from 1950 to 1962 are West Germany 7.3, Italy 6.0, France 4.9, Netherlands 4.7, Denmark 3.5, Norway 3.5, United States 3.3, Belgium 3.2, United Kingdom 2.3. In several countries growth rates were quite different in the two subperiods. In some, 1950-55 was quite disturbed by recovery from World War II.

My method of analyzing growth distinguishes broadly between the contribution of changes in inputs and the contribution of changes in output per unit of input. To measure the contribution of labor, capital, and land the change in each of these inputs, subdivided as necessary and possible among components, must first be measured. The growth rate of each input is then multiplied by its share of national income to obtain its contribution to the growth rate of national income.[2] The contribution of all inputs together is the growth rate of total factor input when the separate inputs are combined by use of income share weights. The contribution of output per unit of input is the amount of the growth rate that is not explained by the growth of inputs. To divide the contribution of output per unit of input among its sources required a separate technique for each source. Broadly speaking, I have tried to isolate the contributions made by important changes in resource allocation, by economies of scale, and by a few sources, such as the change in the average age of capital and the balancing of the capital stock in the early years in Germany, that were either not very important or im-

---

[1] Edward F. Denison, *The Sources of Economic Growth in the United States and the Alternatives before Us* (C.E.D., 1962).

[2] A different procedure was adopted for dwellings and international assets.

portant only in a particular country. I have also tried to isolate the
effect on growth rates of output per unit of input of differences be-
tween the years compared with respect to the pressure of demand upon
employed resources and, insofar as it affected farm output, of weather.
A couple of adjustments were also made for differences in deflation
techniques. I was left with a residual representing the contribution of
advances in knowledge, any catching-up of technique, the contribution
of all changes not explicitly measured such as in how hard people
work, and, of course, noncompensating errors in the growth rates
themselves and in the estimates for the sources dealt with explicitly.

Since the Econometric Society is a joint sponsor of this session, I
had better state that my methodology makes almost no use of correla-
tion analysis. It is quantitative, however, and thus qualifies as econo-
metric according to the charter of the Society.

To measure labor input I took account, sequentially, of changes in
employment, changes in hours worked annually by employed persons,
changes in the distribution of man-hours among individuals classified
by age and sex, and changes in the composition of weighted man-hours
classified by amount of education of the individuals working. I shall
say a few words about each.

The United States had the second largest employment increase
among the nine countries from 1950 to 1962. Germany had the largest.
If there had been no change in the composition of employment, its in-
crease would have contributed 1.5 percentage points to the growth rate
in Germany, 0.9 in the United States, 0.8 to 0.4 in the Netherlands,
Denmark, the United Kingdom, Italy, and Belgium, and as little as 0.1
in France and Norway.

Annual hours worked by full-time nonfarm wage and salary workers
dropped least—almost not at all—in the United States and France.
They dropped most, about 13 percent, in Germany, and by intermedi-
ate amounts elsewhere. But in the United States much of the increase
in employment consisted of women and students voluntarily working
part time, most of them very few hours a week. This pulled down the
average hours of all workers. In Italy, in contrast, involuntary part-
time employment was reduced as job opportunities expanded. Else-
where changes in part-time employment were minor. I have estimated
the net effect on the work done in a year of changes in full-time hours,
an assumed partial productivity offset, and changes in the importance
of part-time employment. It subtracted 0.2 from the United States
growth rate and about as much in five of the eight European countries.
The subtraction was larger in Germany. It was negligible in France
and the contribution was slightly positive in Italy.

The effect of changes in age-sex composition upon average labor

quality was measured by use of hourly earnings weights. The change was most unfavorable in the United States where the share of female man-hours rose most. I estimate it subtracted 0.1 from the United States growth rate. In several countries the contribution was positive —as much as plus 0.1 in France and Italy.

Education is a very important aspect of labor quality. In the United States the education of the labor force has been rising rapidly for several decades. In the 1950-62 period it was rising more in the United States than in any of the European countries and much more than in most of them. Based on a modification of the technique used in my previous study, I estimate that the increased education of the labor force raised the average quality of labor enough to contribute 0.5 percentage points to the United States growth rate from 1950 to 1962. The amount was 0.4 in Belgium and Italy, 0.3 in France and the United Kingdom, 0.2 in the Netherlands and Norway, and only 0.1 in Denmark and Germany. These estimates are based on the use of different weights for the various education levels in the United States and Europe. The United States margin over Europe in the increase in education as such is even greater than these numbers suggest.

When all four aspects are combined, I find that the increase in labor input contributed 1.1 percentage points to the United States growth rate and that only in Germany, where employment increased much more, was the contribution of labor input larger than this.

Capital was divided among four components in this study. The contribution of dwellings to the growth rate can be measured directly by turning to the details of the national accounts to obtain the net value of housing services. The contribution was greatest, 0.25, in the United States. Germany followed with 0.14. The contribution of international assets can also be measured directly. It was 0.05 in the United States, and this amount was exceeded only in the Netherlands. Increases in the stocks of nonresidential structures and equipment and of inventories contributed 0.5 to growth in the United States and this was less than in the European countries except Belgium and the United Kingdom. Germany obtained by far the largest contribution from these sources, 1.4.

All types of capital together contributed 0.8 to the 1950-62 growth rate in the United States and about the same amount in the European countries as a group. The contribution was much larger than this, 1.4, in Germany, moderately higher in the Netherlands and Denmark than in the United States, and appreciably lower in Belgium and the United Kingdom.

The quantity of land was estimated not to have changed significantly, and land therefore made no contribution to growth.

If growth depended only on increases in inputs and if all countries operated under constant returns to scale, 1950-62 growth rates would have been as follows: Germany 2.8, United States 2.0, Netherlands 1.9, Italy 1.7, Denmark 1.6, France and Belgium 1.2, United Kingdom 1.1, and Norway 1.0.

The most pervasive reason that there is so little correspondence between a ranking of countries by growth rates of national income and by growth rates of factor input in this period is to be found, according to my analysis, in the reallocation of resources. I tried to deal with three aspects: the contraction of agriculture, the contraction of nonfarm self-employment, and the reduction of barriers to international trade.

All the countries had in 1950, and still have, a larger proportion of total resources, but mainly people, in agriculture than the proportion that would provide a maximum national income. From 1950 to 1962 the percentage of employment devoted to agriculture declined by 30 to 47 percent in each of the nine countries. These are enormous declines to occur in only twelve years. The percentage decline was big everywhere but the importance of agricultural employment, and therefore the effect upon nonfarm employment, varied greatly. In 1950 agricultural employment ranged from 5 percent of total employment in the United Kingdom, 11 in Belgium, and 12 in the United States to 25 to 29 in Germany, Denmark, and France and 43 in Italy.

National income per unit of input is generally much lower in agriculture than in nonfarm industries. In addition, given time for adjustment, the addition of inputs to nonagricultural industries could raise nonagricultural national income, excluding income from dwellings and abroad, about proportionately, whereas agriculture was so overmanned that elimination of labor in agriculture had only a small adverse effect on agricultural output. My estimates of the contribution to the 1950-62 growth rate that was obtained by shifting agricultural resources to nonfarm activities range from less than 0.1 in the United Kingdom and 0.2 in the United States and Belgium to 0.7 in France, 0.8 in Germany, and 1.0 in Italy.

There is overallocation of labor to nonfarm self-employment that is generally similar to the situation in agriculture. Nonfarm proprietors and unpaid family workers represent very different proportions of total nonfarm employment in different countries, and the proportion in each of the nine countries declined from 1950 to 1962. This group, of course, includes a core of professionals, proprietors of substantial establishments, repairmen, and the like, who do well enough operating as proprietors and in no sense are misallocated. But it is not this group who account for differences between places and dates but a fringe

group, usually much larger, who are very inefficiently employed or underemployed. As large numbers were eliminated and transferred to wage and salary employment, the work they did could be replaced by those remaining or by a disproportionately small increase in paid employment in larger establishments. Gains from this transfer, I estimate, ranged from 0.04 percentage points in the United States and United Kingdom to 0.22 to 0.26 in Italy, France, Norway, and the Netherlands.

My estimates of gains from reductions in barriers to international trade, which are very crude, ranged from nothing in the United States and 0.02 in the United Kingdom to 0.15 or 0.16 in Belgium, the Netherlands, Norway and Italy.

The combined contributions made to 1950-62 growth rates by these three aspects of resource reallocation were as follows: United Kingdom 0.1, United States 0.3, Belgium 0.5, Netherlands 0.6, Denmark 0.7, Norway 0.9, France and Germany 1.0, and Italy 1.4. The differences are big.

In the 1955-62 subperiod the contributions of inputs and resource allocation together pretty well established the ranking of the nine countries by growth rates, aside from the effects on output per unit of input of incomparability of the years 1955 and 1962 with respect to demand pressures and weather. In France and perhaps Italy, however, there appears to have been an added ingredient. To discuss it requires reference to the residuals I obtained after deducting from growth rates the contributions made by all sources of growth explicitly estimated. In the United States itself the residual contributed 0.76 in both 1950-55 and 1955-62. After minor adjustments for procedural comparability, my former study yields about the same amount for the period since the 1920's. I think this residual for the United States can be used as a tolerable estimate of the contribution made to growth by advances in knowledge. In 1955-62 I obtained residuals in the narrow range of 0.75 to 0.97 for seven of the nine countries. In addition to the United States they were Belgium, Denmark, the Netherlands, Germany, the United Kingdom, and Norway. Since these are residuals, I do not regard the small differences among these countries as significant. For France, however, I obtained a residual of 1.56 and for Italy of 1.30. The French residual also exceeded that for the United States by about the same amount in 1950-55. I am inclined to believe that the large French residuals indicate that France was obtaining an appreciable contribution to growth from sources I have not isolated—perhaps a catching up of technological and managerial techniques toward the United States level as a result of deliberate efforts and the pressure of increased competition, elimination of redundancy in nonfarm wage and

salary employment, improvement of resource allocation in ways I have not measured, better incentives and harder work, or perhaps something else.

In the 1950-55 period, Germany and, to a lesser extent, Italy were apparently securing large increases in output per unit of input from the elimination of wartime distortions. I tried to estimate the effects of one or two aspects of this, but for the most part my techniques are inadequate to deal with so disturbed a situation and it is reflected in large residuals for Germany and Italy in 1950-55.[3]

My general conclusion thus far is that the ranking of countries with respect to growth rates in the 1950-62 period as a whole was determined by changes in inputs, gains from resource reallocation, some catching up of technique or other unisolated source in France and possibly Italy, and recovery from war distortions in the early postwar years in Germany and Italy.

The size of the differentials among countries in growth rates was, however, widened by economies of scale. To some extent this was simply because gains from economies of scale depend on the rate at which the size of markets increase, so that the more national income rises for other reasons the greater are gains from scale economies. There is, however, a much more important reason, one that I rather despair of describing even sketchily in the time available.

I have estimated the amount by which European growth rates would be reduced if the components of consumption were reweighted by use of United States rather than European prices. For 1950-62 as a whole the amounts range from 0.1 in Belgium, Norway, and the United Kingdom and 0.2 in Denmark and the Netherlands to 0.5 in France, 0.6 in Italy, and 0.9 in Germany. The reason rates are reduced is that, for different products, the ratio of the volume of European to United States consumption is lower and the ratio of European to United States prices is higher the greater is the income elasticity for a product. As European per capita consumption has risen, the increases have been concentrated in products that have high income elasticity and high relative prices in Europe as compared with the United States. The larger the rise in per capita consumption the larger is this dispersion. The greater the increase in per capita consumption and the lower the level of per capita consumption the more a shift to United States weights reduces the growth rate. I used an indirect method of estimate that picks up only this systematic tendency but checked the results against those obtained by direct reweighting.

If you prefer, you can consider the estimates just given as simply a

---

[3] The residual in 1950-55 is also particularly large in the Netherlands and small in Denmark. I have no special explanation.

statistical adjustment required to place all the national income estimates on a common basis with respect to consumption price weights. I believe, however, that it is predominantly an economies-of-scale phenomenon. As per capita consumption in the fast growing European countries increased, increases were concentrated in products produced in particularly small volume and at particularly high unit cost in Europe as compared with the United States. Techniques of large-scale production were already available in the United States and could be adopted as soon as markets developed to sufficient size.

In addition to sources of growth, I have tried to isolate the sources of difference in levels of national income per person employed. Measured in United States prices, national income per person employed in all the European countries covered except Italy was 58 to 65 percent of the United States level in 1960. In Italy it was 40 percent.

I now offer two groups of observations that emerge from a comparison of sources of growth with sources of difference in income levels.

My first comments concern resource allocation. France and Germany had a much greater potential to raise national income by increasing nonfarm wage and salary employment at the expense of agricultural and self-employment than did the United States, and this provides an important part of my explanation of why output per unit of input rose more in France and Germany. Greater overallocation is also one reason the level of income is lower in France and Germany than in the United States but it is a relatively minor reason. France and Germany are exhausting this source of growth very rapidly, and its complete elimination would go only a small way to narrow their income gap with the United States.

Reallocation of resources also provides much of the explanation of why French and German growth rates were above the British rate. But the level of national income per person employed was as high in the United Kingdom as in France and Germany in 1960 only because the United Kingdom had much less misallocation. In nonfarm industries alone output per person was appreciably lower in the United Kingdom and, indeed, not greatly above Italy. Reduction of misallocation is opening an income gap between France and Germany, on the one hand, and the United Kingdom, and closing the gap between the United Kingdom and Italy.

Agricultural and self-employment depressed the Italian income level most, and this was a principal reason income in Italy was below the other European countries. Their eventual contraction to the same proportions as in the other countries would eliminate much of the gap in income levels. However, education and capital also contribute to the gap.

My second comment concerns what I can only describe as residual productivity. My estimates imply that if there were no difference in the quantity and quality of labor per worker, the amount of capital and land per worker, the size of markets, the cost of misallocation of resources of the types mentioned, or the pressure of demand upon resources, residual productivity in the European countries as a group, except Italy, would still have been 28 percent below the United States in 1960. Whatever other changes may occur, European income per worker cannot approach the American level unless this gap is greatly reduced. My estimates of growth sources indicate that up to 1962, at least, the gap was not being cut much except in France, and even in France the amount seems small compared to the apparent potential.

The United States in 1925 had already reached the level of national income per person employed that was attained by the European countries except Italy, as a group, in 1960. One striking statistic emerging from my study is that residual productivity was lower in Europe in 1960 than it was in the United States in 1925. The gain in the United States from 1925 to 1960 was mainly due, I believe, to advances in knowledge that accumulated over a 35-year period. But disparities in the availability of knowledge as to how to produce cannot explain the 1960 gap between the United States and Europe. One can hardly avoid speculating as to what is responsible, but I shall not do so today.

I shall close with one final generalization. My study indicates that the continental countries did not obtain higher growth rates than we after 1950 because they were doing more to obtain growth. We were drawing more additional workers, particularly women, into employment than any of the other countries. The education of our labor force increased more. Only we and the French did not reduce the hours of full-time workers significantly. We did the most research and development. We cut down farm employment and fringe nonfarm self-employment as much, proportionately, as the European countries. We did not increase the capital stock as much, in percentage terms, as most of the European countries but this situation is less clear than it may seem, and in any case only Germany obtained much more growth than we from this source. The reasons that European growth rates were higher in the postwar period must be sought primarily in differences in initial conditions.

In conclusion, let me state that I keenly regret that I have been unable to provide here the evidence for my conclusions or to introduce the cautions that are usual and appropriate. A book will remedy these omissions in the near future.[4]

[4] Edward F. Denison with the assistance of Jean-Pierre Poulliet, *Why Growth Rates Differ: Postwar Experience in Nine Western Countries* (to be published by the Brookings Institution, 1967).

DISCUSSION

HENRY ROSOVSKY: It is very difficult to disagree with the "lessons" which Professor Cameron finds in the historical record. History, he tells us, demonstrates the virtues of greater efficiency, hard work, education, and the spirit of private enterprise. But all this seems to me somewhat obvious and much too general. If I were a development planner, a bureaucrat, or a businessman in what is here euphemistically called a "developing nation," I would thank the economic historian with courtesy and sincerity, and then send a cable (probably collect) to Messrs. Chenery and Tinbergen saying: "Please come quickly." I say this, because it seems to me that Professor Cameron's lessons, at least by implication, concentrate more on the attributes of economic backwardness than on ways to achieve economic growth.

If we want to draw operational policy implications from the history of modern economic growth, there are at least three prerequisites:

1. We must have one or more explicit theories of the industrialization process.

2. We must, if the lessons are to be useful, restrict the scope of our observations rather severely.

3. We must also, in order to be honest, accentuate the negative.

Let me now say something about each one of these points as they pertain to Professor Cameron's observations.

I find in Professor Cameron's paper too little attention paid to the changing historical nature of the industrialization process. He tells us that "initial conditions" vary, and that Marx's dictum concerning the "mirror image" is unacceptable. And yet his theory concerning the influence of changing initial conditions is never made explicit, and many of his lessons could lead one to believe that Marx was, after all, correct. (Something which I do not believe.) Take the author's remarks concerning capital. He finds that in advanced countries, during the early phases of industrialization, the proportion of net capital formation to national product was surprisingly low, and that these rates rose only after modern economic growth was well under way. "The big development in getting the growth process under way was a more intensive and efficient use of both the existing stock of capital and of normal increments to that stock." This may have been true when the comparatively rich countries of the world industrialized in the nineteenth century. Since then, however, both the nature of technology and the income levels of countries desiring to industrialize have changed. The more efficient use of existing capital stock may have too little impact in parts of Africa, and "normal increments," whatever that may mean, could be far too slow. Underdeveloped countries must import technology, and much of it is capital intensive. It does little good to remind them that this was less true fifty or seventy-five years ago.

This brings me to the need for restricting the scope of observations. Professor Cameron's developed countries run all the way from the United States— by way of Europe—to Japan. The rest of the world is largely underdeveloped.

Yet the developed and the underdeveloped parts of the world are even less homogeneous than the factor capital. For example, there are—historically and currently—young and relatively empty countries and older and settled countries. In the past, the rates of savings in these countries differed in accordance with this classification, and perhaps the lessons should take these differences into account. (The United States always had a surprisingly high NDCF/ NNP ratio. Could this have some useful implications for some countries today?) As things stand, Professor Cameron's conclusions tend to overgeneralize. He says: "The kinds of men to be found in the service of the state—and more importantly the kinds of incentives that impel them—are not necessarily those most conducive to economic growth. . . . It is a common observation that state-sponsored development projects tend towards the grandiose and monumental, frequently of dubious economic value. . . ." If we leave out the qualifying clauses, as we must (the kinds of men to be found anywhere are not necessarily this or that), these statements are far too broad to be useful. The bureaucrats of Meiji Japan and of early nineteenth century Prussia did a lot for economic growth and avoided the monumental. The same may be true of Tito's Yugoslavia. Let me add that I do not think that proves anything one way or the other. My own feeling is that for historical lessons to be helpful we must force ourselves to operate at a much lower level of abstraction. I believe that we could learn a great deal by studying regional patterns of development; i.e., the historical process by which poorer regions of developing countries were drawn into the national market, and the accompanying declines in regional income differentials. How Japan exploited the growth potential inherent in sharp regional variation may contain valuable suggestions for South Korea and Thailand. Beyond that I would not go, but I would remain very satisfied if the assumptions are verified.

My last point has to do with the accentuation of the negative. In our reading of the historical record we all too frequently seem rather too satisfied with ourselves, and we must tend to give the impression that "our way" was the best and most alternatives are quite impossible. But Germany was a leading European aggressor twice in this century; Russia went through a great revolution; Japan tried to conquer much of Asia in the 1930's. These events may suggest some of the disequilibrating influences, or let us say the pathology, of modern economic growth, and they are also part of our history. To me, at least, they illustrate some of the dangers of "growth at any cost," and they deserve full consideration when giving advice to others.

We have listened today to an interesting sermon by a reformed socialist. His present state of mind seems to be rather Weberian and this is as good as any other, because to me, in deducing lessons from the complicated record of history, the most suggestive motto has always been "let a hundred flowers bloom."

CHARLES P. KINDLEBERGER: Today's paper by Mr. Denison is but the tip of a massive iceberg represented by the forthcoming book, *Why Growth Rates Differ,* which he was kind enough to send me in manuscript as background for this comment. It is surely inappropriate for me to review that book before it has appeared, and yet I cannot forebear from saying that just as his

Sources of Growth in the United States and the Alternatives Before Us lifted
the discussion of United States growth to a new level of discourse, so the book
he has written with the assistance of Jean-Pierre Poullier will mark a new
stage in the discussion of European growth and productivity. The profession
will be deeply in his debt for the monumental energy, patience, and informed
judgment he has brought to the measurement of the sources of European
growth from 1950-62 and their comparison with those in the United States.
The work is a model of patient and detailed exposition, admirable candor,
wide-ranging combing of material, and ingenuity of estimation and assump-
tion. If there is room for doubt as to his conclusions, this is the result not of
Mr. Denison and Mr. Poullier's skill being wanting, but of possible limitations
in the method.

There is an apocryphal quotation ascribed to Raymond Goldsmith that
what cannot be measured does not exist. Mr. Denison comes close to this view
at one point where he refers to "noneconomic or unmeasured quality change."
On the whole, however, he is content to measure what he can and speculate in
restrained fashion about what he cannot. But measurement occasionally dis-
torts, as when quality of education is represented by years of schooling, cor-
rected for changes in school days per year and absenteeism of pupils, without
reference to the qualitative distinctions which have long been thought relevant
to growth between the education of an Oxford gentleman and that of the
French or German scientist at the *Grandes Ecoles* or the *Technische Hoch-
schulen,* respectively. And the need to measure comparatively, over a number
of countries, may lose valuable data. In allowing for the gain from realloca-
tion, Denison measures only the movement from the farm and from self-em-
ployed nonfarm workers to industry, and chooses to ignore the improvement
of efficiency from reallocation within sectors and within plants of varying sizes
in manufacturing branches. I believe these effects were of great importance in,
say, France. The sort of contrast which the statistical Procrustean bed misses
is the well-known difference between British railroads with their large num-
bers of workers, thought by some observers to be almost half in disguised un-
employment, and France where the S.N.C.F. since the war has cut down· em-
ployment while building up service in a remarkable fashion.

But the basic question must be directed to method. Can the factors con-
tributing to growth be estimated as if the world's production functions are lin-
ear and homogeneous so that the contribution of separate inputs can be esti-
mated from their income shares; and are the contributions of individual fac-
tors separable? As Mr. Denison's paper states, changes in inputs are first mea-
sured in their contribution to growth, with the remaining contribution of
changes in output per unit of input ingeniously divided among, mainly, scale,
allocation, and knowledge effects, plus a residual which includes such immea-
surables as quality of management, effort, differences in competition, etc.
(The advance of knowledge is in fact part of the residual, but the extension of
the United States estimated residual which is identified as "advance in knowl-
edge" to other countries leaves a still smaller residual for other causes and er-
rors in estimation which do not cancel.)

Are these sources of inputs and of increased efficiency per unit of inputs

separable? Denison recognizes and allows for "interactions" between improved allocation between farm and nonfarm, and from self-employed nonfarm to paid employment. He also frequently notes connections between effects measured separately, such as expansion of labor and economies of scale. But he is not prepared to test the hypothesis that economic growth is a systems problem in which the variables have important interrelations. As a student of trade, for example, I am perhaps unduly moved by the statistical result that the reduction of barriers to international trade produced only 8 hundredths of one percentage point, or 2 percent of the total rate of growth of Northwest Europe in 1950 to 1962. Denison deliberately overlooks the contribution of exports and import-substitution to growth on the macroeconomic or demand level, since his is a supply, not a demand analysis, in sharp distinction to those of Maddison, Lamfalussy, Dow, etc. I applaud the attention to supply, and this point does not trouble me. But increased trade is given credit only for a very small improvement in resource allocation from improved comparative advantage, whereas I should have attributed to it some of the improvement due to better allocation between farm and nonfarm sectors, between self-employed nonfarm and other nonfarm, and the unmeasured improvement in efficiency from reallocation within manufacturing and within manufacturing branches.

I suppose my major concern is that I have just finished a study of European postwar growth—largely intuitive rather than quantitative—which applies the W. Arthur Lewis model of growth with unlimited supplies of labor to various countries in Europe and finds it suggestive. In this way of looking at things, the availability of extra labor, either new to the labor force or reallocated from inefficient sectors, holds down wages, sustains profits, and maintains investment at a high level. When the supply of labor ceases to be elastic, increases in demand lead to higher wages, lower profits, reduced investment, a reversal of the balance of payments, and so on. The various elements which Denison studies carefully and quantitatively, one by one, are fitted into a model. There is difficulty in making it fit all cases, and France and Austria pose special difficulties. But unlike Denison's approach, the economy is viewed as a system rather than as an aggregation.

It is of some interest that Mr. Denison's studies end in 1962 just as Europe ran out of the large supplies of labor which had sustained growth from 1950 or so. On my showing, the supergrowth of the 1950's will not be repeated after 1962 or 1963 or 1964. Mr. Denison's analysis may lead to the same conclusion but there is no way of knowing it. I can illustrate my methodological point best, perhaps, by referring to a footnote of Lionel Robbins to the effect that the only impact of the Great Depression on the National Bureau was to make them change their trend lines, in contradistinction to Keynes who used it to find a new insight.

But my major comment is the one with which I started. Mr. Denison and his assistant have given us numbers for analysis and food for thought to occupy us for years to come. It is a superb achievement. Mr. Denison has isolated most and perhaps the most important sources of observed growth in quantitative form. The task is now to reassemble them into a coherent whole.

PROPERTY RIGHTS AND BEHAVIOR

STOCK VERSUS MUTUAL SAVINGS AND LOAN ASSOCIATIONS: SOME EVIDENCE OF DIFFERENCES IN BEHAVIOR*

By ALFRED NICOLS

University of California, Los Angeles

The "mutual" savings and loan association originated as a self-help cooperative in which each saver received one "share" for every $100 of savings up to a maximum of fifty shares.[1] Though ultimate power is today supposed to reside in the shareholders, the latter typically surrender their proxies to management on opening an account. The consequence has been that while the fifty-share maximum has effectively diffused ownership, control has been concentrated in management. Subject to little, if any, pressure from savers, management is a self-perpetuating autocracy. A West Virginia court even held it permissible to pay off a dissenting shareholder who had asked for a list of shareholders in order to conduct a proxy fight. Not surprisingly, one legal authority concluded that "a shareholder has a vote until he cares to exercise it, and then he ceases to have a vote, if the board so decides."[2] Lenin's retort to Bertrand Russell's complaint that there were no political parties in the Soviet Union is relevant. "A bourgeois lie. Russia has many parties: the communist party in power; and the rest in prison." It is significant that it was not an official of an eastern European "people's democracy" but the executive director of the U.S. Cooperative League[3] who rationalized that it was necessary to protect the "beneficial ownership . . . of the depositors—not only now but yesterday and tomorrow."[4] Today's mutual savings and loan association has more in common with the monolithic Soviet dictatorship than with the

[1] Each borrower is permitted one share.

[2] P. James Riordan, Hearings before House of Representatives, Committee on Banking and Currency (Subcommittee on Bank Supervision and Insurance), 85th Cong., *Federal Charter.* Legislation for Mutual Savings Banks, p. 206.

[3] *Ibid.,* p. 296, Testimony, Jerry Voorhis.

[4] It has been observed by Arthur H. Courshon, National League of Insured Savings and Loan Associations: "The current Board has taken the position that the shareholders in a mutual institution are represented by the Board. They are not represented by themselves." *Hearings,* p. 251.

benevolent and democratic cooperative societies of nineteenth-century America.

The only constraint is that management is limited to a wage which must be "reasonable." On the other hand, management has not only been allowed but encouraged to establish privately-owned ancillary corporations which provide services to both the customers and the association: mortgage brokerage, insurance, title, trustee. It is the purpose of this paper to examine both the mechanism by which management converts its power position into personal gain and the evidence relating to the efficiency of the mutual as compared to the stock savings and loan association. Performance is measured in terms of growth in assets, market share changes, expenses, gross operating income, and allocations to reserves.

I. Market Structure in the Savings and Loan Industry

The savings and loan industry is similar to Caesar's Gaul in that it is divided into three parts. Associations operating under federal charter, all mutuals, had $56.4 billion in assets at the end of 1963, 52.5 percent of the total assets of the Federal Home Loan Bank system. State chartered mutuals had $29.8 billion in assets, 27.7 percent of the total. State chartered savings and loan associations of the "stock" type have increased by 436 percent or from $4.0 billion in 1955 to $21.2 billion, a rise in market share from 11.0 percent to 19.7 percent. California stock associations constituted 62 percent of total stock assets. Stock associations are restricted to twenty states.

II. A Model for a Mutual Savings and Loan Association

At the outset it should be noted that the stress is on the differences between the two types of ownership and control. It is not argued that stock associations are completely run and dominated by their stockholders. Stockholders are interested in earnings; and at times exert pressure on management; and are often successful in disposing of that management. Furthermore, even when stockholders are relatively passive there is always the possibility, however remote, that outsiders will make a raid on the stock of the association.

Management of mutuals are free from such headaches and are proud that their decisions need not show earnings designed to inflate stock prices. They boast that their mutual is not "up for grabs" by the wrong people.[5] Presumably the "right" people are presently in control.

The problem of self-perpetuating management in a mutual is one of conversion. How is managerial power converted into personal wealth? There are choices but each is limited:

[5] *Ibid.* Testimony, J. E. Hoeft, Chairman of the Board and President, Glendale Savings & Loan Association of Glendale, California, pp. 289-90; Adolf A. Berle, pp. 307, 308.

1. Though income may be maximized in the form of compensation or wages, it is not possible to directly convert all the profits. Growth in the size of the savings and loan association in order to increase compensation involves certain disadvantages. First, risk exposure is increased and loss of control where insolvency is a probable result is threatened. Since management only receives the benefits of that growth to a limited extent, it would merely add to its worries. Furthermore, there is an asymmetry between success and failure in a mutual. Management receives only part of the rewards of success and the entire blame for the failure!

2. The management may exercise power in the form of amenities taken on the association's premises; e.g., attractive furnishings, beautiful secretaries, "agreeable" persons reflecting the same ethnic and religious background, association-owned automobiles, lengthy vacations with pay, and large expense accounts. But even amenities are subject to check: (a) the association must work within the constraint that revenues cover all costs if it is to remain solvent; (b) the Internal Revenue Service is quick to identify amenities as payments in lieu of income for which taxes must be paid.

3. The diversion of potential association income into wholly owned ancillary establishments. Management in a mutual savings and loan association is a power unto itself. It determines both output and the prices charged both for the services rendered and the prices paid suppliers.

Consider a disequilibrium situation, where there are only mutual associations in each market and entry is restricted by the regulatory authority. Associations in growth periods would receive prices in excess of operating costs. Where a stock association, operating under normal profit maximizing incentives expands because marginal revenue is in excess of marginal costs, the mutual need not and would not. By not doing so its behavior is similar in nature to a restrictive monopolist. A reluctance to move towards profit equilibrium is expected of all mutuals regardless of the market situation. This relative insensitivity of mutual output to profit establishes a margin greater than that for the profit maximizing equilibrium and is as permanent as is the disequilibrium from which it is derived.

But not only is there less reason for the mutual to offer the equilibrium output, there is also no reason why the market would be permitted to determine prices. Where under conventional profit maximization reduced output would imply an excess demand, or "shortage," which would force price up, the mutual sets price on the basis of expenses. His optimal price is considerably less than one which "clears" the market.

The conversion mechanism rests on the manager's power over out-

put and price. Diversion of income to his affiliates requires permanent disequilibrium or restriction coupled with prices below those which would clear the market. Given the shortage, management sets nominal prices and bargains with individuals over the premium they must pay. The price of a loan includes more than the interest rate and fees but also the purchase of related services. The stock association with no necessity for diversion will produce a larger volume and its prices will be smaller by the value to the borrower from securing the related services in free or more competitive markets.

The mutual management thereby "rations" loans on the basis of private deals with borrowers. For example, he may be more willing to lend to borrowers who purchase insurance from his private insurance company. The borrower gains a loan he would not otherwise receive. The value to the mutual is illustrated: in 1964, had the largest five California mutuals secured only one-third of the insurance on their outstanding loans, each would have picked up an estimated $129,585 additional income. In 1964, 4.13 percent of the assets of federally chartered associations in Los Angeles were held in the form of cash on hand or in commercial banks. For the five largest federals with an average $642 million in assets, each averaged $27 million. With certificates of deposit selling at 4.0 percent, the return would have been $1,000,000. If the management of a large federal decided to forego interest[6] in favor of a demand deposit, it is obvious that a bank would be most appreciative.

The placement of savings and loan association cash flows is not inconsiderable. When "loans repaid" are added to net savings for 1964, the five largest federals would have had about a $148 million flow and the next five $56 million. Commercial banks can create many opportunities for acknowledging appreciation. Lines of credit readily become available to the savings and loan association officials, and may be used to purchase options for speculative land appreciation—an arrangement not permitted officers of the savings and loan associations with their own institution. Or relatives and business associates in other fields, on the reference of the savings and loan official, may secure loans for which they would otherwise not qualify. Whether it be commercial bank deposits or the sale of insurance, escrow, or trustee relationships is immaterial; management is in a position to use the margin between prices and cost in a bargaining arrangement which is similar to a "tie-in" sale.

[6] The New York legislative inquiry of 1905 (The Armstrong Committee) revealed that the large New York insurance companies had maintained unduly large cash balances in banks in which both the officers and their companies were heavily involved. (Joint Committee of the Senate and Assembly, New York, Investigation of Life Insurance Companies, Report, Vol. X, pp. 30-31, 83-84, 132).

The mutual association will not grow as rapidly as the stock association, but will also have less incentive for competitive increases in the dividend rates. Competition at any level is to be avoided because it reduces the margin and thereby destroys the mechanism for diversion. Without the margin, the management has nothing to offer the customer or supplier in return for receipt of a line of credit at the bank or the insurance contract. Mutuality restricts the dimensions of competition in that the management cannot afford to forego the tie-in arrangement. Accordingly, the chances for price competition among mutuals are that much smaller—whatever the initial degree of competition or oligopoly may have been. Greater reliance is placed on ancillary activities.

III. *The Evidence*

The hypothesis derived from the above model is that mutual savings and loan associations, in contrast to stock savings and loan associations, will report a retarded rate of growth associated both with lower gross operating income and lower dividend rates paid to savers. The relevant comparisons are between the California federal and stock associations and the California stock associations as compared to the remainder of the United States (which is dominated by the mutual form). The reason for the two comparisons is that the California federally chartered associations have had to develop some of the attributes of the stock associations in order to survive. Unlike mutuals elsewhere in the nation, they must compete with stock associations. Consequently, the differences between California federals and California stock would not be as large as that between the latter and mutuals outside of California.

Expenses. Emphasis in a mutual is on expenses rather than profit; and is directed to both managerial compensation as well as the indirect and nonpecuniary benefits of the manager's position. When California federal associations are compared to stock associations in 1962, the latter with a new loan volume of $3.8 million was twice the federals with $1.8 million. Yet the stock associations' expenses at $119.3 million were only 43 percent larger than the federals' expenses at $73.8 million (Table 1). On the basis of a variant of Parkinson's Law, expenses are what income permits.

Conventional measurement of expenses on assets or gross operating income is unsatisfactory. Variations in product mix, e.g., the ratio of "loans made" to assets, or the "construction" loan to asset ratio, influence the expense ratio and, therefore, make comparisons misleading. The trouble is that both assets and gross operating income are derived from assets acquired in previous years. Once placed on the books such assets make significantly smaller additions to expenses.

TABLE 1

OPERATING RESULTS: U.S. COMPARED TO CALIFORNIA,
MUTUAL AND STOCK ASSOCIATIONS
1962

	New Loans÷ Average Assets	Expenses ÷Average Assets	Gross Operating Income÷ Average Assets	Fees— Loans Made	Dividends ÷Average Saving Capital	Allocation to Reserve and Sur- plus÷ Average Net Worth
U.S. less California insured members FHLB..........	21.3	1.14	5.55	1.49	4.01	12.6
California federals..	31.8	1.28	6.38	2.32	4.59	18.1
California stock....	41.3	1.31	7.10	3.14	4.84	22.0
1963						
U.S. less California.	21.2	1.17	5.48	1.29	4.04	9.3
California federals..	34.9	1.36	6.31	2.14	4.72	11.8
California stock....	45.1	1.42	6.60	2.77	4.78	15.4

SOURCE: Federal Home Loan Bank Board, *Combined Financial Statements*, 1961–63, and *Source Books*, 1962–64 (Washington, D.C.).

In order to escape from these difficulties, ratios were abandoned in favor of a multiple regression analysis using actual data with respect to forty-eight of the states and thirty-five[7] of the largest standard metropolitan statistical areas (SMSA's). (See Table 2.) The major explanatory variables were dollar volume of "loans made," "new savings," and the average number of mortgage accounts. Loans made were further subdivided into "construction" loans, "other" loans, and "home purchase."

The regression analysis indicated that if loans made had been doubled in 1962 as compared to the mean for the series, expenses would have been increased by 65.5 percent for the states and 59.1 percent for the SMSA's. When the California and Los Angeles stock data with respect to these variables is related to the regression coefficients, a mutual area with the same product mix would have had expenses of 62.6 percent and 72.6 percent higher than what was actually achieved by the stock associations.[8]

[7] California was excluded because of the influence of the stock associations. Alaska was both minor and isolated. Five cities were excluded because of the dominant influence of stock associations: Los Angeles, Phoenix, San Bernardino, San Jose, and San Francisco.

[8] When allowance is made for estimated greater average size of "loans made" by California institutions, the excess of the mutuals is reduced to 30 percent. But this adjustment is not thought necessary for three reasons: (1) when size of loan is tested as an independent variable, it turned out to be statistically insignificant; (2) it is not necessary to make an adjustment when comparisons are made within the same market area, e.g., California where average size of loans made is roughly identical for both mutuals and stock: (3) an adjust-

TABLE 2

OPERATING COST REGRESSION MODELS

	Constant Term	Loans Made	Number of Mortgages	New Savings	R^2
35 SMSA's 1962.....	−0.33062				.9850
Reg. Coef..........		0.03433	−0.00514	0.01479	
(s.e.)..............		(.00423)	(.01158)	(.00262)	
Part. Cor. Coef......		0.82429	−.07947	0.71134	
35 SMSA's 1963.....	−1.47961				.9893
Reg. Coef..........		0.03059	0.05646	0.00735	
(s.e.)..............		(.00562)	(.01347)	(.00236)	
Part. Cor. Coef......		0.69930	0.60139	0.48794	
35 SMSA's 1964.....	−1.46038				.9893
Reg. Coef..........		0.02461	0.06718	0.00812	
(s.e.)..............		(.00648)	(.01725)	(.00384)	
Part. Cor. Coef......		0.56337	0.57320	0.35510	
48 States 1962 (Mutual only)......	−659				.9888
Reg. Coef..........		0.03551	0.00868	0.01041	
(s.e.)..............		(.00353)	(.00554)	(.00178)	
Part. Cor. Coef......		0.83491	0.22985	0.66117	

SOURCE: Federal Home Loan Bank Board, *Combined Financial States*, 1962–64, and *Savings and Home Financing: Source Book*, 1963–64.

Trend in Expenses, 1950 to 1964. When the ratio of total operating expenses is averaged for the years 1962-64 and compared to the average for 1950-52, the United States, without California, showed an increase of 434 percent. (See Table 3.) This is compared to an increase in activity as measured by loans made of 285 percent. The number of mortgage accounts increased by 122 percent; and the number of savings accounts, 186 percent. Since the major item in expenses is loan origination, it is significant that the increase in expenses was 52 percent higher than the increase in loans made. California stock expenses increased 1,296 percent, while "loans made" rose 1,363 percent.[9] The ratio of incremental expenses to incremental loans was .95 as compared to 1.52 for the United States without California. Where expenses increased by 434 percent for the U.S., assets increased 457 percent. In contrast, the

---

ment on the basis of "average size" does not solve problems: it only changes them because average size is a ratio of dollar volume of mortgages to number of mortgages. Mortgages are heterogeneous, differing as to expense involved in placing on the books: riskier loans often entail more work. Therefore, it seemed preferable to eschew tampering with the lid of this Pandora's box. Even in a competitive market, the stock generated twice as much new loan volume with but 43 percent higher expenses. On the basis of 1960 census information indicating differences in size of loans made, California stock data were deflated by 23.8 percent. The difference was only reduced to 41.6 percent.

[9] California stock association data for 1950-52 estimated on the basis of asset and loans made shares.

California stock associations' expenses increased 1,296 percent and assets 1,702 percent.[10]

  Volume of Operations. It has been suggested above that measurement by total assets is misleading since the volume of assets in no way reflects the activity in any one period. To illustrate, an association of $100 million in assets may "originate" $20 million in loans a year, while another association of similar asset size or even less might "originate" $40 million in loans. Obviously, the latter association will have higher expenses.

    Limitations of space preclude presentation of data with respect to volume for stock associations as compared to federals in the eleven states where both operate and for which there is data. Suffice it to say,

TABLE 3

TRENDS IN EXPENSES AS RELATED TO OTHER VARIABLES, 1950–64
% Changes

	Expenses	Loans Made	Assets	Number of Mortgages	Number of Savings Accounts
U.S. less California.	434	285	457	122	186
California federals..	499	390	609	277	336
California stocks...	1,296	1,363	1,702	554	1,039

SOURCE: Federal Home Loan Bank Board, *Combined Financial Statements*, 1950–64.

in the years 1960-62, the stock associations' ratio of loans made to average mortgages was higher in nine states for each of the years. In addition, this greater ratio of loans made also brought the stock associations' higher gross operating and net income as related to average assets. (See Table 1.)

  Market Share Changes. For California, the market share of the federally chartered associations fell from 60.6 percent of total assets in 1950 to 33.7 percent in 1964. This is consistent with the hypothesis that the mutual associations will show a retarded rate of growth.

  Changes in Net Worth and Assets for California Associations, 1953-62 (Existing December 31, 1953): Federal Compared to Stock. For associations in existence at the end of 1953 (see Table 4), the federals increased assets by 323 percent. The stock associations increased 640 percent. But more significant is the increase in net worth since this

  [10] The comparison here repeats the possible shortcoming indicated in note 8: the relatively higher increase in percentage of loans made by California stock associations could be associated with a greater increase in average size of loans made. On the basis of estimated changes in average mortgage size 1950-64, U.S. (less California) was 82.1 percent of California. California stock loans made were deflated with the result that the increase in expenses came to 115.8 percent of the increase in new loans. Notwithstanding the adjustment, the U.S. increase was 31.5 percent more than this deflated estimate.

represents interest-free earning assets. The federal associations' net worth increased $319.6 million, or 327 percent. The stock associations increased $607.9 million, or 649 percent.

Nepotism. It is expected that since nepotism is both a nonpecuniary reward and a tax-free form of inheritance, it is larger in mutuals. Nepotism is defined as a relationship of the chief executive officer to other officers. The relationship is simply the same last name and obviously understates the degree of nepotism because it fails to include in-laws or relatives with different surnames. In 1955, the stock associations reported 27 incidents for 146 members, or a ratio of 18 percent; the federally chartered, 22 incidents for 66 members, or 33 percent.[11]

TABLE 4

CHANGES IN NET WORTH AND ASSETS FOR CALIFORNIA ASSOCIATIONS EXISTING
ON DECEMBER 31, 1953: FEDERAL COMPARED TO STOCK*
(000's)

	1953	1962	Absolute Change	% Change
Net worth† federal.......	97,676	417,243	319,567	327.
stock.........	93,633	701,567	607,934	649.
Total assets federal.......	1,440,404	6,097,108	4,656,704	323.
stock.........	1,234,591	9,141,204	7,906,613	640.

* *1963 Roster of Members,* Federal Home Loan Bank of San Francisco; California Savings and Loan Commissioner, *69th Annual Report.*
† *Combined Financial Statements,* 1953 and 1962, Federal Home Loan and Bank Board.

For 1963, the stock showed 15 percent and the federally chartered, 43 percent.

Management Changes, 1950-64.[12] Change in management is defined as a change in the chairman, president, or listed manager, or any combination thereof. Expressing the management changes as a percentage of the number of associations, the median value reported for the stock associations was 18.9 percent; for the federal associations, 5.9 percent.

Competition for Savings. The California stock associations made greater use of advertising media than did the federals: 59 compared to 12 federals.[13] They also stressed interest rate where the federals referred to "ratio of reserve to savings." When federals are compared to stock within the same states, the ratio of dividends paid to average savings capital was higher for the stock for the years 1960-62. Stock associations were also more effective merchandisers, as revealed in their development and aggressive use of premiums and bonuses. In

[11] Federal Home Loan Bank of San Francisco, *Roster of Members,* 1955 and 1963.
[12] *Ibid.,* 1950-64.
[13] Data derived from *New York Times,* 1955-63, first Sunday edition of each quarter; and similar method used for Los Angeles *Times.*

short, the federal have, on the average, competed only most reluctantly.[14]

IV. *Significance of the Findings*

If mutuality is characterized by public benovolence and stock ownership private cupidity, it might be better if the nation opted for the latter. Mutuality is a euphemism which has effectively shielded powerful monopolistic forces from the public scrutiny. Furthermore, if California experience is any guide, the mutual is an institution which can survive only at the cost of continued governmental restriction of competition; i.e., through achieving a "sacred cow" status similar to agriculture.

The remedy for the excessive margins of mutuality is obvious. If mutuals were converted into stock associations, not only would profit provide a basis for rational decision making, but the necessity for maintenance of nonequilibrium margins would vanish. Removal of the profit constraint on management eliminates the pressure to "second best" choices such as nepotism, infrequent management changes, and tie-in "deals"—all of which increase the price to borrowers while reducing the interest paid to savers. The diversion mechanism is not only restrictive, it is inefficient.

[14] Not only have the stock associations paid higher rates, they initiated the increases last summer when Home Savings and Loan of Los Angeles, the nation's largest, went to 5¼, 0.40 more than the Federal Home Loan Bank set for member associations requesting advances.

TOWARD A THEORY OF PROPERTY RIGHTS

By HAROLD DEMSETZ
University of Chicago

When a transaction is concluded in the marketplace, two bundles of property rights are exchanged. A bundle of rights often attaches to a physical commodity or service, but it is the value of the rights that determines the value of what is exchanged. Questions addressed to the emergence and mix of the components of the bundle of rights are prior to those commonly asked by economists. Economists usually take the bundle of property rights as a datum and ask for an explanation of the forces determining the price and the number of units of a good to which these rights attach.

In this paper, I seek to fashion some of the elements of an economic theory of property rights. The paper is organized into three parts. The first part discusses briefly the concept and role of property rights in social systems. The second part offers some guidance for investigating the emergence of property rights. The third part sets forth some principles relevant to the coalescing of property rights into particular bundles and to the determination of the ownership structure that will be associated with these bundles.

The Concept and Role of Property Rights

In the world of Robinson Crusoe property rights play no role. Property rights are an instrument of society and derive their significance from the fact that they help a man form those expectations which he can reasonably hold in his dealings with others. These expectations find expression in the laws, customs, and mores of a society. An owner of property rights possesses the consent of fellowmen to allow him to act in particular ways. An owner expects the community to prevent others from interfering with his actions, provided that these actions are not prohibited in the specifications of his rights.

It is important to note that property rights convey the right to benefit or harm oneself or others. Harming a competitor by producing superior products may be permitted, while shooting him may not. A man may be permitted to benefit himself by shooting an intruder but be prohibited from selling below a price floor. It is clear, then, that property rights specify how persons may be benefited and harmed, and, therefore, who must pay whom to modify the actions taken by persons. The recognition of this leads easily to the close relationship between property rights and externalities.

Externality is an ambiguous concept. For the purposes of this paper, the concept includes external costs, external benefits, and pecuniary as well as nonpecuniary externalities. No harmful or beneficial effect is external to the world. Some person or persons always suffer or enjoy these effects. What converts a harmful or beneficial effect into an externality is that the cost of bringing the effect to bear on the decisions of one or more of the interacting persons is too high to make it worthwhile, and this is what the term shall mean here. "Internalizing" such effects refers to a process, usually a change in property rights, that enables these effects to bear (in greater degree) on all interacting persons.

A primary function of property rights is that of guiding incentives to achieve a greater internalization of externalities. Every cost and benefit associated with social interdependencies is a potential externality. One condition is necessary to make costs and benefits externalities. The cost of a transaction in the rights between the parties (internalization) must exceed the gains from internalization. In general, transacting cost can be large relative to gains because of "natural" difficulties in trading or they can be large because of legal reasons. In a lawful society the prohibition of voluntary negotiations makes the cost of transacting infinite. Some costs and benefits are not taken into account by users of resources whenever externalities exist, but allowing transactions increases the degree to which internalization takes place. For example, it might be thought that a firm which uses slave labor will not recognize all the costs of its activities, since it can have its slave labor by paying subsistence wages only. This will not be true if negotiations are permitted, for the slaves can offer to the firm a payment for their freedom based on the expected return to them of being free men. The cost of slavery can thus be internalized in the calculations of the firm. The transition from serf to free man in feudal Europe is an example of this process.

Perhaps one of the most significant cases of externalities is the extensive use of the military draft. The taxpayer benefits by not paying the full cost of staffing the armed services. The costs which he escapes are the additional sums that would be needed to acquire men voluntarily for the services or those sums that would be offered as payment by draftees to taxpayers in order to be exempted. With either voluntary recruitment, the "buy-him-in" system, or with a "let-him-buy-his-way-out" system, the full cost of recruitment would be brought to bear on taxpayers. It has always seemed incredible to me that so many economists can recognize an externality when they see smoke but not when they see the draft. The familiar smoke example is one in which negotiation costs may be too high (because of the large number of interact-

ing parties) to make it worthwhile to internalize all the effects of smoke. The draft is an externality caused by forbidding negotiation.

The role of property rights in the internalization of externalities can be made clear within the context of the above examples. A law which establishes the right of a person to his freedom would necessitate a payment on the part of a firm or of the taxpayer sufficient to cover the cost of using that person's labor if his services are to be obtained. The costs of labor thus become internalized in the firm's or taxpayer's decisions. Alternatively, a law which gives the firm or the taxpayer clear title to slave labor would necessitate that the slaveowners take into account the sums that slaves are willing to pay for their freedom. These costs thus become internalized in decisions although wealth is distributed differently in the two cases. All that is needed for internalization in either case is ownership which includes the right of sale. It is the prohibition of a property right adjustment, the prohibition of the establishment of an ownership title that can thenceforth be exchanged, that precludes the internalization of external costs and benefits.

There are two striking implications of this process that are true in a world of zero transaction costs. The output mix that results when the exchange of property rights is allowed is efficient and the mix is independent of who is assigned ownership (except that different wealth distributions may result in different demands).[1] For example, the efficient mix of civilians and military will result from transferable ownership no matter whether taxpayers must hire military volunteers or whether draftees must pay taxpayers to be excused from service. For taxpayers will hire only those military (under the "buy-him-in" property right system) who would not pay to be exempted (under the "let-him-buy-his-way-out" system). The highest bidder under the "let-him-buy-his-way-out" property right system would be precisely the last to volunteer under a "buy-him-in" system.[2]

We will refer back to some of these points later. But for now,

[1] These implications are derived by R. H. Coase, "The Problem of Social Cost," *J. of Law and Econ.*, Oct., 1960, pp. 1-44.

[2] If the demand for civilian life is unaffected by wealth redistribution, the assertion made is correct as it stands. However, when a change is made from a "buy-him-in" system to a "let-him-buy-his-way-out" system, the resulting redistribution of wealth away from draftees may significantly affect their demand for civilian life; the validity of the assertion then requires a compensating wealth change. A compensating wealth change will not be required in the ordinary case of profit maximizing firms. Consider the farmer-rancher example mentioned by Coase. Society may give the farmer the right to grow corn unmolested by cattle or it may give the rancher the right to allow his cattle to stray. Contrary to the Coase example, let us suppose that if the farmer is given the right, he just breaks even; i.e., with the right to be compensated for corn damage, the farmer's land is marginal. If the right is transferred to the rancher, the farmer, not enjoying any economic rent, will not have the wherewithal to pay the rancher to reduce the number of head of cattle raised. In this case, however, it will be profitable for the rancher to buy the farm, thus merging cattle raising with farming. His self-interest will then lead him to take account of the effect of cattle on corn.

enough groundwork has been laid to facilitate the discussion of the next two parts of this paper.

The Emergence of Property Rights

If the main allocative function of property rights is the internalization of beneficial and harmful effects, then the emergence of property rights can be understood best by their association with the emergence of new or different beneficial and harmful effects.

Changes in knowledge result in changes in production functions, market values, and aspirations. New techniques, new ways of doing the same things, and doing new things—all invoke harmful and beneficial effects to which society has not been accustomed. It is my thesis in this part of the paper that the emergence of new property rights takes place in response to the desires of the interacting persons for adjustment to new benefit-cost possibilities.

The thesis can be restated in a slightly different fashion: property rights develop to internalize externalities when the gains of internalization become larger than the cost of internalization. Increased internalization, in the main, results from changes in economic values, changes which stem from the development of new technology and the opening of new markets, changes to which old property rights are poorly attuned. A proper interpretation of this assertion requires that account be taken of a community's preferences for private ownership. Some communities will have less well-developed private ownership systems and more highly developed state ownership systems. But, given a community's tastes in this regard, the emergence of new private or state-owned property rights will be in response to changes in technology and relative prices.

I do not mean to assert or to deny that the adjustments in property rights which take place need be the result of a conscious endeavor to cope with new externality problems. These adjustments have arisen in Western societies largely as a result of gradual changes in social mores and in common law precedents. At each step of this adjustment process, it is unlikely that externalities per se were consciously related to the issue being resolved. These legal and moral experiments may be hit-and-miss procedures to some extent but in a society that weights the achievement of efficiency heavily, their viability in the long run will depend on how well they modify behavior to accommodate to the externalities associated with important changes in technology or market values.

A rigorous test of this assertion will require extensive and detailed empirical work. A broad range of examples can be cited that are consistent with it: the development of air rights, renters' rights, rules for

liability in automobile accidents, etc. In this part of the discussion, I shall present one group of such examples in some detail. They deal with the development of private property rights in land among American Indians. These examples are broad ranging and come fairly close to what can be called convincing evidence in the field of anthropology.

The question of private ownership of land among aboriginals has held a fascination for anthropologists. It has been one of the intellectual battlegrounds in the attempt to assess the "true nature" of man unconstrained by the "artificialities" of civilization. In the process of carrying on this debate, information has been uncovered that bears directly on the thesis with which we are now concerned. What appears to be accepted as a classic treatment and a high point of this debate is Eleanor Leacock's memoir on *The Montagnes "Hunting Territory" and the Fur Trade.*[3] Leacock's research followed that of Frank G. Speck[4] who had discovered that the Indians of the Labrador Peninsula had a long-established tradition of property in land. This finding was at odds with what was known about the Indians of the American Southwest and it prompted Leacock's study of the Montagnes who inhabited large regions around Quebec.

Leacock clearly established the fact that a close relationship existed, both historically and geographically, between the development of private rights in land and the development of the commercial fur trade. The factual basis of this correlation has gone unchallenged. However, to my knowledge, no theory relating privacy of land to the fur trade has yet been articulated. The factual material uncovered by Speck and Leacock fits the thesis of this paper well, and in doing so, it reveals clearly the role played by property right adjustments in taking account of what economists have often cited as an example of an externality— the overhunting of game.

Because of the lack of control over hunting by others, it is in no person's interest to invest in increasing or maintaing the stock of game. Overly intensive hunting takes place. Thus a successful hunt is viewed as imposing external costs on subsequent hunters—costs that are not taken into account fully in the determination of the extent of hunting and of animal husbandry.

Before the fur trade became established, hunting was carried on primarily for purposes of food and the relatively few furs that were required for the hunter's family. The externality was clearly present. Hunting could be practiced freely and was carried on without assessing its impact on other hunters. But these external effects were of such

---

[3] Eleanor Leacock, *American Anthropologist* (American Anthropological Asso.), Vol. 56, No. 5, Part 2, Memoir No. 78.

[4] Cf., Frank G. Speck, "The Basis of American Indian Ownership of Land," *Old Penn Weekly Rev.* (Univ. of Pennsylvania), Jan. 16, 1915, pp. 491-95.

small significance that it did not pay for anyone to take them into account. There did not exist anything resembling private ownership in land. And in the *Jesuit Relations,* particularly Le Jeune's record of the winter he spent with the Montagnes in 1633-34 and in the brief account given by Father Druilletes in 1647-48, Leacock finds no evidence of private land holdings. Both accounts indicate a socioeconomic organization in which private rights to land are not well developed.

We may safely surmise that the advent of the fur trade had two immediate consequences. First, the value of furs to the Indians was increased considerably. Second, and as a result, the scale of hunting activity rose sharply. Both consequences must have increased considerably the importance of the externalities associated with free hunting. The property right system began to change, and it changed specifically in the direction required to take account of the economic effects made important by the fur trade. The geographical or distributional evidence collected by Leacock indicates an unmistakable correlation between early centers of fur trade and the oldest and most complete development of the private hunting territory.

> By the beginning of the eighteenth century, we begin to have clear evidence that territorial hunting and trapping arrangements by individual families were developing in the area around Quebec. . . . The earliest references to such arrangements in this region indicates a purely temporary allotment of hunting territories. They [Algonkians and Iroquois] divide themselves into several bands in order to hunt more efficiently. It was their custom . . . to appropriate pieces of land about two leagues square for each group to hunt exclusively. Ownership of beaver houses, however, had already become established, and when discovered, they were marked. A starving Indian could kill and eat another's beaver if he left the fur and the tail.[5]

The next step toward the hunting territory was probably a seasonal allotment system. An anonymous account written in 1723 states that the "principle of the Indians is to mark off the hunting ground selected by them by blazing the trees with their crests so that they may never encroach on each other. . . . By the middle of the century these allotted territories were relatively stabilized."[6]

The principle that associates property right changes with the emergence of new and reevaluation of old harmful and beneficial effects suggests in this instance that the fur trade made it economic to encourage the husbanding of fur-bearing animals. Husbanding requires the ability to prevent poaching and this, in turn, suggests that socioeconomic changes in property in hunting land will take place. The chain of reasoning is consistent with the evidence cited above. Is it inconsistent with the absence of similar rights in property among the southwestern Indians?

Two factors suggest that the thesis is consistent with the absence of

[5] Eleanor Leacock, *op. cit.,* p. 15.
[6] Eleanor Leacock, *op. cit.,* p. 15.

similar rights among the Indians of the southwestern plains. The first of these is that there were no plains animals of commercial importance comparable to the fur-bearing animals of the forest, at least not until cattle arrived with Europeans. The second factor is that animals of the plains are primarily grazing species whose habit is to wander over wide tracts of land. The value of establishing boundaries to private hunting territories is thus reduced by the relatively high cost of preventing the animals from moving to adjacent parcels. Hence both the value and cost of establishing private hunting lands in the Southwest are such that we would expect little development along these lines. The externality was just not worth taking into account.

The lands of the Labrador Peninsula shelter forest animals whose habits are considerably different from those of the plains. Forest animals confine their territories to relatively small areas, so that the cost of internalizing the effects of husbanding these animals is considerably reduced. This reduced cost, together with the higher commercial value of fur-bearing forest animals, made it productive to establish private hunting lands. Frank G. Speck finds that family proprietorship among the Indians of the Peninsula included retaliation against trespass. Animal resources were husbanded. Sometimes conservation practices were carried on extensively. Family hunting territories were divided into quarters. Each year the family hunted in a different quarter in rotation, leaving a tract in the center as a sort of bank, not to be hunted over unless forced to do so by a shortage in the regular tract.

To conclude our excursion into the phenomenon of private rights in land among the American Indians, we note one further piece of corroborating evidence. Among the Indians of the Northwest, highly developed private family rights to hunting lands had also emerged— rights which went so far as to include inheritance. Here again we find that forest animals predominate and that the West Coast was frequently visited by sailing schooners whose primary purpose was trading in furs.[7]

[7] The thesis is consistent with the development of other types of private rights. Among wandering primitive peoples the cost of policing property is relatively low for highly portable objects. The owning family can protect such objects while carrying on its daily activities. If these objects are also very useful, property rights should appear frequently, so as to internalize the benefits and costs of their use. It is generally true among most primitive communities that weapons and household utensils, such as pottery, are regarded as private property. Both types of articles are portable and both require an investment of time to produce. Among agriculturally-oriented peoples, because of the relative fixity of their location, portability has a smaller role to play in the determination of property. The distinction is most clearly seen by comparing property in land among the most primitive of these societies, where crop rotation and simple fertilization techniques are unknown, or where land fertility is extremely poor, with property in land among primitive peoples who are more knowledgeable in these matters or who possess very superior land. Once a crop is grown by the more primitive agricultural societies, it is necessary for them to abandon the land for several years to restore productivity. Property rights in land among such people would require policing cost for several years during which no sizable output is obtained. Since to provide for

The Coalescence and Ownership of Property Rights

I have argued that property rights arise when it becomes economic for those affected by externalities to internalize benefits and costs. But I have not yet examined the forces which will govern the particular form of right ownership. Several idealized forms of ownership must be distinguished at the outset. These are communal ownership, private ownership, and state ownership.

By communal ownership, I shall mean a right which can be exercised by all members of the community. Frequently the rights to till and to hunt the land have been communally owned. The right to walk a city sidewalk is communally owned. Communal ownership means that the community denies to the state or to individual citizens the right to interfere with any person's exercise of communally-owned rights. Private ownership implies that the community recognizes the right of the owner to exclude others from exercising the owner's private rights. State ownership implies that the state may exclude anyone from the use of a right as long as the state follows accepted political procedures for determining who may not use state-owned property. I shall not examine in detail the alternative of state ownership. The object of the analysis which follows is to discern some broad principles governing the development of property rights in communities oriented to private property.

It will be best to begin by considering a particularly useful example that focuses our attention on the problem of land ownership. Suppose that land is communally owned. Every person has the right to hunt, till, or mine the land. This form of ownership fails to concentrate the cost associated with any person's exercise of his communal right on that person. If a person seeks to maximize the value of his communal rights, he will tend to overhunt and overwork the land because some of the costs of his doing so are borne by others. The stock of game and the richness of the soil will be diminished too quickly. It is conceivable that those who own these rights, i.e., every member of the community, can agree to curtail the rate at which they work the lands if negotiating and policing costs are zero. Each can agree to abridge his rights. It is obvious that the costs of reaching such an agreement will not be zero. What is not obvious is just how large these costs may be.

Negotiating costs will be large because it is difficult for many per-

---

sustenance these people must move to new land, a property right to be of value to them must be associated with a portable object. Among these people it is common to find property rights to the crops, which, after harvest, are portable, but not to the land. The more advanced agriculturally based primitive societies are able to remain with particular land for longer periods, and here we generally observe property rights to the land as well as to the crops.

sons to reach a mutually satisfactory agreement, especially when each hold-out has the right to work the land as fast as he pleases. But, even if an agreement among all can be reached, we must yet take account of the costs of policing the agreement, and these may be large, also. After such an agreement is reached, no one will privately own the right to work the land; all can work the land but at an agreed upon shorter workweek. Negotiating costs are increased even further because it is not possible under this system to bring the full expected benefits and expected costs of future generations to bear on current users.

If a single person owns land, he will attempt to maximize its present value by taking into account alternative future time streams of benefits and costs and selecting that one which he believes will maximize the present value of his privately-owned land rights. We all know that this means that he will attempt to take into account the supply and demand conditions that he thinks will exist after his death. It is very difficult to see how the existing communal owners can reach an agreement that takes account of these costs.

In effect, an owner of a private right to use land acts as a broker whose wealth depends on how well he takes into account the competing claims of the present and the future. But with communal rights there is no broker, and the claims of the present generation will be given an uneconomically large weight in determining the intensity with which the land is worked. Future generations might desire to pay present generations enough to change the present intensity of land usage. But they have no living agent to place their claims on the market. Under a communal property system, should a living person pay others to reduce the rate at which they work the land, he would not gain anything of value for his efforts. Communal property means that future generations must speak for themselves. No one has yet estimated the costs of carrying on such a conversation.

The land ownership example confronts us immediately with a great disadvantage of communal property. The effects of a person's activities on his neighbors and on subsequent generations will not be taken into account fully. Communal property results in great externalities. The full costs of the activities of an owner of a communal property right are not borne directly by him, nor can they be called to his attention easily by the willingness of others to pay him an appropriate sum. Communal property rules out a "pay-to-use-the-property" system and high negotiation and policing costs make ineffective a "pay-him-not-to-use-the-property" system.

The state, the courts, or the leaders of the community could attempt to internalize the external costs resulting from communal property by allowing private parcels owned by small groups of person with similar

interests. The logical groups in terms of similar interests, are, of course, the family and the individual. Continuing with our use of the land ownership example, let us initially distribute private titles to land randomly among existing individuals and, further, let the extent of land included in each title be randomly determined.

The resulting private ownership of land will internalize many of the external costs associated with communal ownership, for now an owner, by virtue of his power to exclude others, can generally count on realizing the rewards associated with husbanding the game and increasing the fertility of his land. This concentration of benefits and costs on owners creates incentives to utilize resources more efficiently.

But we have yet to contend with externalities. Under the communal property system the maximization of the value of communal property rights will take place without regard to many costs, because the owner of a communal right cannot exclude others from enjoying the fruits of his efforts and because negotiation costs are too high for all to agree jointly on optimal behavior. The development of private rights permits the owner to economize on the use of those resources from which he has the right to exclude others. Much internalization is accomplished in this way. But the owner of private rights to one parcel does not himself own the rights to the parcel of another private sector. Since he cannot exclude others from their private rights to land, he has no direct incentive (in the absence of negotiations) to economize in the use of his land in a way that takes into account the effects he produces on the land rights of others. If he constructs a dam on his land, he has no direct incentive to take into account the lower water levels produced on his neighbor's land.

This is exactly the same kind of externality that we encountered with communal property rights, but it is present to a lesser degree. Whereas no one had an incentive to store water on any land under the communal system, private owners now can take into account directly those benefits and costs to their land that accompany water storage. But the effects on the land of others will not be taken into account directly.

The partial concentration of benefits and costs that accompany private ownership is only part of the advantage this system offers. The other part, and perhaps the most important, has escaped our notice. The cost of negotiating over the remaining externalities will be reduced greatly. Communal property rights allow anyone to use the land. Under this system it becomes necessary for all to reach an agreement on land use. But the externalities that accompany private ownership of property do not affect all owners, and, generally speaking, it will be necessary for only a few to reach an agreement that takes these effects into account. The cost of negotiating an internalization of these effects

is thereby reduced considerably. The point is important enough to elucidate.

Suppose an owner of a communal land right, in the process of plowing a parcel of land, observes a second communal owner constructing a dam on adjacent land. The farmer prefers to have the stream as it is, and so he asks the engineer to stop his construction. The engineer says, "Pay me to stop." The farmer replies, "I will be happy to pay you, but what can you guarantee in return?" The engineer answers, "I can guarantee you that I will not continue constructing the dam, but I cannot guarantee that another engineer will not take up the task because this is communal property; I have no right to exclude him." What would be a simple negotiation between two persons under a private property arrangement turns out to be a rather complex negotiation between the farmer and everyone else. This is the basic explanation, I believe, for the preponderance of single rather than multiple owners of property. Indeed, an increase in the number of owners is an increase in the communality of property and leads, generally, to an increase in the cost of internalizing.

The reduction in negotiating cost that accompanies the private right to exclude others allows most externalities to be internalized at rather low cost. Those that are not are associated with activities that generate external effects impinging upon many people. The soot from smoke affects many homeowners, none of whom is willing to pay enough to the factory to get its owner to reduce smoke output. All homeowners together might be willing to pay enough, but the cost of their getting together may be enough to discourage effective market bargaining. The negotiating problem is compounded even more if the smoke comes not from a single smoke stack but from an industrial district. In such cases, it may be too costly to internalize effects through the marketplace.

Returning to our land ownership paradigm, we recall that land was distributed in randomly sized parcels to randomly selected owners. These owners now negotiate among themselves to internalize any remaining externalities. Two market options are open to the negotiators. The first is simply to try to reach a contractual agreement among owners that directly deals with the external effects at issue. The second option is for some owners to buy out others, thus changing the parcel size owned. Which option is selected will depend on which is cheaper. We have here a standard economic problem of optimal scale. If there exist constant returns to scale in the ownership of different sized parcels, it will be largely a matter of indifference between outright purchase and contractual agreement if only a single, easy-to-police, contractual agreement will internalize the externality. But, if there are several externalities, so that several such contracts will need to be negotiated, or

if the contractual agreements should be difficult to police, then outright purchase will be the preferred course of action.

The greater are diseconomies of scale to land ownership the more will contractual arrangement be used by the interacting neighbors to settle these differences. Negotiating and policing costs will be compared to costs that depend on the scale of ownership, and parcels of land will tend to be owned in sizes which minimize the sum of these costs.[8]

The interplay of scale economies, negotiating cost, externalities, and the modification of property rights can be seen in the most notable "exception" to the assertion that ownership tends to be an individual affair: the publicly-held corporation. I assume that significant economies of scale in the operation of large corporations is a fact and, also, that large requirements for equity capital can be satisfied more cheaply by acquiring the capital from many purchasers of equity shares. While economies of scale in operating these enterprises exist, economies of scale in the provision of capital do not. Hence, it becomes, desirable for many "owners" to form a joint-stock company.

But if all owners participate in each decision that needs to be made by such a company, the scale economies of operating the company will be overcome quickly by high negotiating cost. Hence a delegation of authority for most decisions takes place and, for most of these, a small management group becomes the *de facto* owners. Effective ownership, i.e., effective control of property, is thus legally concentrated in management's hands. This is the first legal modification, and it takes place in recognition of the high negotiating costs that would otherwise obtain.

The structure of ownership, however, creates some externality difficulties under the law of partnership. If the corporation should fail, partnership law commits each shareholder to meet the debts of the corporation up to the limits of his financial ability. Thus, managerial *de facto* ownership can have considerable external effects on shareholders. Should property rights remain unmodified, this externality would make it exceedingly difficult for entrepreneurs to acquire equity capital from wealthy individuals. (Although these individuals have recourse to reimbursements from other shareholders, litigation costs will be high.) A second legal modification, limited liability, has taken place to reduce the effect of this externality.[9] *De facto* management ownership and limited liability combine to minimize the overall cost of operating large enterprises. Shareholders are essentially lenders of equity capital and not owners, although they do participate in such infrequent decisions as

---

[8] Compare this with the similar rationale given by R. H. Coase to explain the firm in "The Nature of the Firm," *Economica*, New Series, 1937, pp. 386-405.

[9] Henry G. Manne discusses this point in a forthcoming book about the American corporate system.

those involving mergers. What shareholders really own are their shares and not the corporation. Ownership in the sense of control again becomes a largely individual affair. The shareholders own their shares, and the president of the corporation and possibly a few other top executives control the corporation.

To further ease the impact of management decisions on shareholders, that is, to minimize the impact of externalities under this ownership form, a further legal modification of rights is required. Unlike partnership law, a shareholder may sell his interest without first obtaining the permission of fellow shareholders or without dissolving the corporation. It thus becomes easy for him to get out if his preferences and those of the management are no longer in harmony. This "escape hatch" is extremely important and has given rise to the organized trading of securities. The increase in harmony between managers and shareholders brought about by exchange and by competing managerial groups helps to minimize the external effects associated with the corporate ownership structure. Finally, limited liability considerably reduces the cost of exchanging shares by making it unnecessary for a purchaser of shares to examine in great detail the liabilities of the corporation and the assets of other shareholders; these liabilities can adversely affect a purchaser only up to the extent of the price per share.

The dual tendencies for ownership to rest with individuals and for the extent of an individual's ownership to accord with the minimization of all costs is clear in the land ownership paradigm. The applicability of this paradigm has been extended to the corporation. But it may not be clear yet how widely applicable this paradigm is. Consider the problems of copyright and patents. If a new idea is freely appropriable by all, if there exist communal rights to new ideas, incentives for developing such ideas will be lacking. The benefits derivable from these ideas will not be concentrated on their originators. If we extend some degree of private rights to the originators, these ideas will come forth at a more rapid pace. But the existence of the private rights does not mean that their effects on the property of others will be directly taken into account. A new idea makes an old one obsolete and another old one more valuable. These effects will not be directly taken into account, but they can be called to the attention of the originator of the new idea through market negotiations. All problems of externalities are closely analogous to those which arise in the land ownership example. The relevant variables are identical.

What I have suggested in this paper is an approach to problems in property rights. But it is more than that. It is also a different way of viewing traditional problems. An elaboration of this approach will, I hope, illuminate a great number of social-economic problems.

ON THE DISTINCTION BETWEEN PUBLIC AND PRIVATE GOODS*

By OTTO A. DAVIS, *Carnegie Institute of Technology*
and ANDREW B. WHINSTON, *Purdue University*

I. *Introduction*

Some years ago, in a now classic series of articles [13] [14] [15], Professor Paul A. Samuelson made an admittedly polar distinction between public and private goods. Briefly, private consumption goods, like bread, must be parceled out among persons with one man getting a loaf more if another gets a loaf less. Thus, if x_{ij} represents the ith person's consumption of the jth private good and X_j represents the quantity available, then $\sum_i x_{ij} = X_j$. A public consumption good, on the other hand, differs in that one man's consumption does not diminish the quantity available for another. Thus, if y_{ik} represents the ith person's consumption of the kth public good and Y_k represents the quantity available, $y_{ik} = Y_k$ for all men. Prime examples are supposed to be outdoor circuses, national defense, and, some may have thought, radio waves and TV signals.

Recently, responding to Minasian's argument [10] that market allocation (via subscription) of television might be a more desirable arrangement, Professor Samuelson [16] claimed that the possession by television of the characteristics of a public good did not constitute evidence either for or against a market arrangement via subscription. There is no doubt but that Samuelson's claim (or admission?) is correct. Aside from television and radio signals, there are many examples of goods which are allocated in the market place and yet which exhibit various degrees of "Samuelsonian publicness." The products of the researcher's efforts, the outpuring of the jukebox and the record player, undercapacity indoor (why be restricted to the outdoors?) circuses and performances of movies and various forms of entertainment, the unhurried and uncrowded viewing of old masters which are owned by a private art collector, etc.—all exhibit aspects of publicness. Olsen [12] claims that the concept can be applied to all large organizations.[1]

[1] But see Wagner's review [18] of Olsen's theme.

When one is faced with such examples, it is only natural to wonder if it is possible to determine in any objective fashion whether a good should be governmentally or privately produced. Samuelson's classic articles seem to provide few clues to an answer. Yet, the problem appears to merit consideration.

II. *Institutional Arrangements and Samuelson's Model*

It is important to understand that, despite the title, Samuelson's model is not directly concerned with public expenditures. Rather, it is a model of market failure. What is shown is that in a system incorporating his "public" consumption goods, the market "fails" in the sense that the necessary conditions for the attainment of Pareto optimality are not automatically satisfied. One should observe two facts here. First, despite the name given to the public goods, the failure of the pricing mechanism to satisfy the necessary conditions for Pareto optimality does not constitute a prima facie case for public ownership, or even public regulation, of the relevant facilities for either production or distribution. Second, the institutional arrangements under which a good, even a public consumption good, is distributed in a market perform an important role in determining the characteristics of the performance of the pricing mechanism.

Recall the definition of Samuelson's public good. One of the implicit notions is that whatever is produced is available to all. Obviously, even if the appropriate maximization problem yields the necessary optimality conditions with the associated Lagrange multipliers, it is difficult to even imagine real prices performing the role indicated by the multipliers. Behaviorally, the problem appears to be much more complicated and rich than is indicated by the simple mathematics. At issue is the fact that paying the price does not give one control of the good. Instead, the act of paying the price is somehow separated from the act of consumption. Consider the following example in order to clarify the behavioral implications of this point.

Imagine that the government decided to alter the institutional arrangements or organization of the market for bread, which is certainly a private good. Suppose that the government decided to separate the act of paying from the act of obtaining the bread. Accordingly, imagine that payments for the desired quantity of bread were made by consumers to the government in the early morning at the "revenue center." Later in the day, bread could be obtained at the "distribution center." However, suppose that there is no communication between the revenue and distribution centers so that the acts of paying and obtaining are truly separated. The bread is distributed on a first come, first served basis. One might observe that this institutional arrangement would certainly

affect the functioning of the market mechanism. Consumers could not be counted upon to reveal their preferences at the revenue center. Nor would the distribution center fare much better since consumers could simply take the bread they wanted, if it happened to be available, no matter what their payment at the revenue center. Price could not perform its traditional functions. The institutional arrangements cause market failure.

Observe that, at least under one interpretation, there are certain similarities between the above example and the Samuelsonian model of public consumption goods. By definition, a public good in Samuelson's model is available to all consumers no matter whether a consumer chooses to pay or not. From the standpoint of an individual consumer, a payment could not give one control over the public good in the sense of deciding the precise quantity to be consumed. The act of paying is separated from the act of consuming. It is not surprising that the market fails. For a consumer actually to make a payment would be nothing more or less than an act of pure charity.

III. *Some Conditions Needed for the Operation of a Market*

Since he was interested in the problem of attaining an efficient or Pareto optimal allocation, Samuelson did not discuss the possibility of making a market for public goods operate much as if it were allocating private goods. Nevertheless, such a possibility is often available as an alternative to the collective provision of the public good if the appropriate institutional arrangements can be accomplished. Accordingly, a brief and informal discussion of some of the conditions which seem to be needed if a market is to operate in a reasonable manner appears to be merited.

It is obvious that the first requirement is a provision for some type of ownership or property rights. Without ownership, at least in some form, there would seem to be no basis for exchange. Of course, there are complex issues involved in the definition of property rights. The very definition can affect the manner in which exchange takes place and the way in which a market operates.[2] The definition of property rights can sometimes determine whether externalities are allowed to exist in the market. Zoning ordinances, for example, can be interpreted as an effort to remove the influence of externalities in the urban property market.[3] It might be noted that in a sense a good may be characterized by the control an owner may exercise, so that the same piece of property is in effect a different good once the zoning ordinance is changed.

The second requirement, which is related to the issues of ownership

[2] This basic point seems to have been first pointed out by Coase [4] [5]. See also Buchanan and Stubblebine [3] and Davis and Whinston [7].
[3] See Davis [6].

and property rights, concerns "control" over the good or service. The act of buying or paying for the good or service must be related to the use or consumption of it. Even in the case of public good, purchase can sometimes be made to give some control and establish a connection between the act of paying and consumption as the example of television signals with scramblers illustrates.

The third requirement, which is related to both of the previous ones, concerns the possibility of exclusion. While perfect exclusion need not be required for markets to function reasonably well, as numerous examples of the presence of externalities amply illustrates, it is easily seen that exclusion does provide basic motivation for exchange.[4]

IV. *The Market Allocation of Samuelsonian Public Goods*

In part, the question of whether the above requirements can be satisfied for some particular good depends upon the technological characteristics of the good. In order to illustrate this point, consider the case of television. In the age of electronic scramblers, exclusion is possible. The signals can be "owned" and their services sold. Thus it is possible to establish a direct relationship between the consumption of the service and the payment for it. The above requirements for a market can be satisfied. Since the issue does not seem to have been considered fully, it appears appropriate to discuss briefly the operation of such a market on the basis of an excessively simple model.

Consider a society composed of people deciding upon viewing a pay television program. For simplicity, let the capital or fixed costs of producing the program be ignored. Assume that there is no advertising. Consider the following definitions:

$$x_i = \begin{cases} 0 \text{ if the } i\text{th person does not watch the program} \\ 1 \text{ if that person does watch the program} \end{cases}$$

$$y = \begin{cases} 0 \text{ if the program is not produced} \\ 1 \text{ if the program is produced} \end{cases}$$

$u_i(x_i)$: The value or "utility" to the ith person from watching the program.

$g(y)$: A function indicating the amount of some homogeneous (for simplicity) resource used to produce the program.

M: An arbitrarily large number.

K: The "minimal" amount of the homogeneous resource used to produce the program; i.e., $9(1) = K$.

Given these definitions, and assuming that the objective is to attain a

---

[4] Musgrave [11] emphasizes the notion of exclusion.

Pareto optimum in television viewing, consider the following vector maximization problem:

$$(1.1) \quad \max \; [u_1(x_1), \; u_2(x_2), \; \cdots , \; u_n(x_n)]$$

subject to

(1)

$$(1.2) \quad \sum_i x_i - My \leq 0$$

$$(1.3) \quad g(y) \leq K$$

$$(1.4) \quad x_i = 0, 1; \; y = 0, 1 \qquad i = 1, \cdots, n$$

It is convenient to replace the vector (1.1). Following the procedure introduced in [8], let α_i represent the reciprocal of the ith individual's marginal utility of income. Define $d_i = u_i(1)$ and $c_i = \alpha_i d_i$. Assuming for simplicity that the α_i are not affected by the decision to watch or not watch the program, one can replace the vector (1.1) by

$$(2) \qquad \sum_i \alpha_i u_i(x_i) = \sum_i \alpha_i d_i x_i = \sum_i c_i x_i$$

and maximize (2) subject to the same constraints so that the problem is reduced to one of integer maximization. Note that the use of M in (1.2) means that once $y = 1$ the constraint is never binding. The program must be available if someone watches it.

Let p represent the price of the homogeneous resource and assume that p is given by a perfect market. Define $k = p \cdot g(1)$ so that $p \cdot g(y) = ky$. Problem (1) can now be replaced by the following equivalent problem:

$$\max \; \sum_i c_i x_i - ky$$

(3)   subject to

$$\sum_i x_i - My \leq 0$$

and the solution to (3) is obvious. If $\sum_i c_i \geq k$ for $c_i \geq 0$, then the program should be presented and all x_i corresponding to $c_i \geq 0$ should be set equal to one. If there is an i such that $c_i < 0$, then that $x_i = 0$. All persons who enjoy the program should be able to view it.

Consider whether the above solution could be achieved by a non-discriminatory pricing device. Let λ denote the price of being able to view the program. Obviously, if no one who wishes to see the program is to be excluded, then the best that the television network can do is to set

$$(4) \qquad\qquad \lambda = \min_i c_i \geq 0$$

so that the total revenue to the network is $\lambda \sum_i x_i$ where the sum is over all i whose $c_i \geq 0$. Two issues are relevant here. First, $\lambda \sum x_i < k$ is a likely outcome so that a subsidy may be required to pay for the costs of the program. The fact that total revenue is less than variable costs (recall that fixed costs are ignored here) does not necessarily indicate that the program is unwarranted. Second, there seems to be no feasible way for the network to determine the appropriate value of λ as indicated in (4). Obviously, the network cannot ask potential viewers to reveal the value of their c_i. One might think of starting with a zero price and raising it by successive increments until some viewer turned off his set; but surely viewers would catch on to this little game and bluff by turning off their set at any positive price so that it would be driven back down to zero. Therefore, it does not appear that pricing, with or without subsidization, can result in the achievement of the solution to (3).

Since the presence of market failure does not imply the existence of a superior nonmarket alternative, consider the "second-best" type of problem where the television program is priced, there is no subsidy and costs must be covered. It seems obvious in these circumstances that the "best of the second best" possibilities is one which allows costs to be covered exactly. In other words, values of both λ and the x_i must be determined such that

$$(5) \qquad \lambda \sum_i x_i - k = 0$$

and the other conditions are also satisfied. It is not always true that such a solution exists or that it is unique. Assuming existence, presumably one could find such a solution by starting with an extremely low price and then raising it by increments until (5) was satisfied. Note that the existence of (5) removes the incentive for bluffing. In other words, in this second-best type of allocation by pricing, there is no problem of individuals emitting false signals or not revealing true preferences. Observe, however, that this second-best solution does impose costs of exclusion upon the society. This cost is given by $\sum c_i$ where the sum is taken over those c_i satisfying the relation $0 \leq c_i \leq \lambda$ where here λ represents the particular value of λ which was selected to satisfy (5).

Note that no motivational premise, such as profit maximization, was introduced as the basis for the selection of the second-best constraint (5). However, it can be argued that in an expanded model where alternate networks compete for viewers of their programs, the very competition will cause the "no profit" condition (5) to be satisfied so that this analysis does provide a basis (at least as a first approximation) for the assessment of the costs of the private provision of television services by the pricing mechanism.

Observe again that the particular pricing arrangement analyzed above depended upon the technological feasibility of scrambling (and de-scrambling) devices. However, even if these devices were not available, market allocations are still possible if certain institutional actions are taken. Law might require that television sets be made in such a manner that a special device is required to tune into any given channel. The purchase of one of these special devices would give the owner of the set the right to view the specified channel as often as desired. Such an arrangement would not markedly differ from the practice of many private swimming clubs, for example, of selling a season pass instead of charging for each trip to the pool. Similarly, channels might be assigned to the networks who might "rent" the right to make sets capable of tuning into the specified channels to the various manufacturers. Obviously, various arrangements are possible. Each possibility has its own operating characteristics and corresponding social cost.

The point here is that technological considerations can determine partially which institutional arrangements are feasible. The feasibility of various institutional arrangements certainly has an influence in the determination of which particular one is to be selected as the most appropriate. Since technology changes over time, one can expect that institutions should be modified accordingly.

V. *Institutional Choice for Samuelsonian Public Goods*

The recent exchange between Minasian and Samuelson, which was concerned with the issue of subscription television, illustrates the fact that the optimality conditions are only of limited value in making an institutional choice.[5] Samuelson [16] emphasizes that there is no presumption that the particular arrangement which does satisfy the necessary conditions is one which produces a Pareto optimal allocation. Sufficient conditions must also be considered. Even more important, one must determine whether there are feasible institutional arrangements which could result in the satisfaction of both sets of conditions. It is obviously a fact that there is nothing inherent in the derivation of either necessary or sufficient conditions for a Pareto optimum which suggests that feasible arrangements for satisfying these conditions exist. It is interesting to note that Samuelson stressed this very point in the first of his series of papers on this topic.[6]

The consequence of a lack of feasible institutional arrangements for satisfying both necessary and sufficient conditions for Pareto optimality in systems which include public goods is the realization that institutional choice involves the comparison of alternative arrangements which are

[5] See Minasian [10] and Samuelson [16].
[6] See Samuelson [13], pp. 388–89.

necessarily nonoptimal in the sense of Pareto. The problem is not to choose from the Pareto optimum positions that one considered ethically desirable. Rather, the problem is to choose from a feasible set of institutional arrangements that particular one which gives the most suitable or "best" allocation of the good under consideration.

It is obvious that in choosing between alternative institutional arrangements the actual operating characteristics of these various arrangements are factors of great importance. As Buchanan has argued so eloquently, it makes little sense to compare operating characteristics of unobtainably ideal arrangements in making an institutional choice.[7] Consequently, if one is choosing, for example, between the alternative institutional arrangements of subscription television, zero pricing and financing by advertising, or zero pricing without advertising with governmental provision of the service, it is not sufficient to argue that subscription television should be ruled out because a nonzero price violates the necessary condition for efficiency, that financing by advertising results in programs appealing to mass taste so that cultural and educational values are overlooked, and that these reasons indicate governmental provision. One cannot simply assume (and be correct) that governmental provision will be "ideal" simply because the other two arrangements are not. One must consider the actual operating characteristics of all alternative arrangements. The problem of institutional choice often involves the comparison of alternative problems of second best which characterize the available possibilities.

For a Samuelsonian public good, there are certain costs associated with private provision and exclusion. When the requirements for a market are satisfied, at the expense of those costs imposed upon the system by exclusion and the operation of a pricing system, one obtains those advantages provided only by markets.[8] Information concerning desires is provided and can be incorporated into decisions concerning supply, quality of the good, etc. It is worthwhile noting that the relative intensity of desires can be at least partially revealed in even an imperfect market system while this kind of information is much more difficult to obtain under any other arrangement. Nevertheless, the costs of exclusion are not to be taken lightly.

Governmental provision of a public good at, say, a zero price, on the other hand, also has problems associated with it. How are the decision-makers to decide, for example, the quantity which should be provided? One can give the easy answer to this question by saying that the political process makes the decision. However, such an answer only evades

---

[7] See Buchanan [1].

[8] The point that markets are costly to operate has been emphasized by Coase [5]. Demsetz [9] also makes this point but with a greater emphasis upon the costs associated with policing the property rights.

the issue. How does the policitican get information which would lead to the proper decision? One must realize that politicians make decisions at least in part upon the basis of political costs which might be approximated in terms of votes. There does not appear to be any reason to suspect that votes are superior indicators of desires in a system where there are many issues and few elections, so that a vote may enjoy alternative interpretations.

There are obvious difficulties in analyzing those problems of second best where the institutional arrangement is governmental provision, or governmental regulation, of the good under consideration. One does not yet know the appropriate behavioral characteristics which should determine the form of the second-best constraints in the models. What are the behavioral rules which guide the decisions and the actions of governmental agencies? It appears that this question must be answered in a satisfactory and useful manner before there will be a very reasonable basis for analyses capable of adequately determining whether given "public" goods should be governmentally produced and distributed or left to the private sector, with or without regulation.

Finally, it should be noted that ethical considerations cannot completely be ignored in institutional choice. Strotz [17] emphasizes the fact that public goods have important distributional implications. Similarly, alternative institutional arrangements for the allocation of public goods have distributional implications.

VI. *The Allocation of Non-Samuelsonian Public Goods*

It has long been recognized that Samuelson's admittedly polar definition of public goods omits much of governmental activity. It can be argued that another class of goods which should be either governmentally provided or regulated is characterized by extreme decreasing costs. This class of goods involves a high, fixed, and negligible marginal cost coupled with the presence of a capacity constraint. Roads and bridges are prime examples. Leaving aside the question of whether these goods "should" be governmentally produced and distributed, and also leaving the question of how the government actually would do it, let us consider the problem of how such a good "should" be produced and its services allocated over time. One needs to have the benchmark of ideal performance in order to either access or influence actual performance.

Consider a good (say a bridge or a road) which must be constructed in an initial period and whose services must be allocated in that period and also T periods in the future. Supposedly this good becomes obsolete and disintegrates at the end of the Tth period. One incurs the high fixed cost in the initial period and only negligible costs thereafter. Two interrelated questions appear. One must choose the capacity (or decide "how much

to consume") in the initial period and also determine an allocation over time.

In order to answer the above questions, let us consider a very simple model in which there are only two goods. One good (say bread) is produced and consumed during each period. It is an ordinary private good. The other good is the one discussed above. In our system it is often (but not necessarily always) a governmental good and for convenience will be termed such here. Consider the following definitions:

$x^i_1(t)$: The amount of the private good consumed by the ith person during period t.

$x^i_2(t)$: The amount of the governmental good consumed by the ith person during period t.

$y_1(t)$: The quantity of the private good produced and available for consumption during period t. For simplicity, assume no storage or carryover between periods.

$y_2(1)$: The quantity of the governmental good produced during the initial period and available for consumption during each of the T periods.

$g(y_1(t))$: An implicit "production possibility" function relating available resources (assumed to be one) to $y_1(t)$ during periods $t=2, \cdots, T$.

$H(y_1(1), y_2(1))$: An implicit "production possibility" function relating available resources (assumed to be one) to the quantities $y_1(1)$, $y_2(1)$ of the two goods which are produced during the initial period.

U_t^i: The assumed concave utility function of the ith individual during period t.

It is convenient to simplify the problem by ignoring side issues and assuming that there are no externalities in the system. Consequently, it is assumed that in the initial period the productions of the two goods are not functionally interrelated so that H is separable and can be written in the form

(6) $$H(y_1(1), y_2(1)) = h_1(y_1(1)) + h_2(y_2(1))$$

and that the decisions of the initial period do not affect the production possibilities of the following periods so that g can be given as indicated. It is also assumed that utility depends only upon consumption in the given period so that there are no functional interdependencies over time. In addition, it is convenient to assume that utility functions can be written in the separable form

(7) $$U_t^i(x^i_1(t), x^i_2(t)) = u^i_{1t}(x^i_1(t)) + u^i_{2t}(x^i_2(t))$$

so that there is no interaction. The utility from bread does not influence the utility from crossing the bridge and vice versa.

Since interest here is centered upon deriving the conditions for Pareto optimality, it is obvious that the problem is one of vector maximization. It is convenient, however, to apply the Kuhn-Tucker equivalence theorem and follow the procedure outlined in [8] in order to state the criterion function in a more appropriate form. Accordingly, and as before, let α_i represent the (unspecified but assumed positive) reciprocal of the ith consumer's marginal utility of income. The maximization problem can be written as follows:

(8.1) $$\max \sum_i \sum_t \alpha_i u_{1t}^i(x_1^i(t)) + \sum_i \sum_t \alpha_i u_{2t}^i(x_2^i(t))$$

   subject to

(8)

(8.2) $$\sum_i x_1^i(t) \leq y_1(t) \qquad t = 1, \cdots, T$$

(8.3) $$\sum_i x_2^i(t) \leq y_2(1) \qquad t = 1, \cdots, T$$

(8.4) $h_1(y_1(1)) + h_2(y_2(1)) \leq 0$

(8.5) $g(y_1(t)) \leq 0 \qquad t = 2, \cdots, T$

(8.6) $x_1^i(t) \geq 0, \; x_2^i(t) \geq 0, \; y_1(t) \geq 0, \; y_2(1) \geq 0 \qquad \begin{array}{l} t = 1, \cdots, T \\ i = 1, \cdots, n \end{array}$

Note that (8.1) is the criterion (social welfare) function which is stated in a form useful for determining the conditions for Pareto optimality. Constraint (8.2) states that no more of the private good can be consumed than is available in any given period. Constraint (8.3) indicates that no more of the governmental good can be consumed in any period than is made available in the first period. Note that the quantity available of the governmental good is measured in terms of capacity. Constraints (8.4) and (8.5) indicate that no more of the goods can be produced than is allowed by the available resources. Finally, (8.5) indicates the nonnegativity conditions.

It is obvious from the very structure of problem (8) that the usual and familiar conditions apply for the private good. Accordingly, these Pareto conditions are not presented here. Competitive markets are fully capable of satisfying these conditions and performing the desirable allocation. Attention is centered on the conditions for the production and allocation of the governmental good. Let $\lambda(t)$ represent the multiplier associated with constraint (8.3). Recall that this multiplier can be interpreted as a shadow price. Then the conditions on the demand side are

$$(9.1) \quad \alpha_i \frac{\partial u_{2t}^i}{\partial x_2^i(t)} - \lambda(t) \begin{Bmatrix} \leq \\ = \end{Bmatrix} 0 \quad \text{if} \quad x_2^i(t) \begin{Bmatrix} = \\ > \end{Bmatrix} 0, \qquad \begin{aligned} i &= 1, \cdots, n \\ t &= 1, \cdots, T \end{aligned}$$

which are also the conventional and familiar Pareto conditions. If the governmental good is chosen, then an individual should consume that amount in each period which equates his weighted marginal utility to the price of the good during the period. Let β represent the multiplier associated with (8.4). Then on the supply side the condition is

$$(9.2) \qquad \sum_t \lambda(t) - \beta \frac{\partial h_2}{\partial y_2(1)} \begin{Bmatrix} \leq \\ = \end{Bmatrix} 0 \quad \text{if} \quad y_2(1) \begin{Bmatrix} = \\ > \end{Bmatrix} 0$$

which has an interesting interpretation. Note that if one were to determine the quantity (or capacity) of the governmental good to be made available according to a profit maximization criterion with the prices taken as given, then one would consider the problem

$$(10) \qquad \max_{y_2(1) \geq 0} \left\{ \sum_t \lambda(t) y_2(1) - \beta h_2(y_2(1)) \right\}$$

and the conditions for the solution to this problem are given by (9.2). Note that condition (9.2) can be interpreted as saying that if the governmental good is to be supplied, then that quantity (capacity) should be chosen in the initial period which will equate the sum of the prices to the marginal cost of supplying the selected quantity. The solution is that given by profit maximization.

Under the usual assumptions, the conditions derived from (8) are both necessary and sufficient for Pareto optimality. Observe that, in regard to the governmental good, one obtains as a solution to (8) a vector of consumption quantities $(x_2^i(1), \cdots, x_2^i(T))$ for each i, a number $y_2(1)$ which is the quantity made available in each period (the capacity), and a vector of prices or charges $(\lambda(1), \cdots, \lambda(T))$. This solution is Pareto optimal. Note, however, that it cannot be decentralized period by period.

There are some interesting aspects to the solution to this problem which have important implications for the planning for the provision of the class of goods (such as bridges and roads) under consideration. Note that if demand is at all variable, it is likely that the charge $\lambda(t)$ which rations the available supply $y_2(1)$ is likely to be zero for some periods. From the point of view of planning, the zero charge does not imply no exclusion. It merely means that a person does not have to pay to obtain the right to use the bridge, but he must still obtain that right. Also note the implication of (9.2). The capacity (quantity supplied) should not be selected to be so large that the corresponding constraint (8.3) never becomes binding. The constraint must become binding during some

periods so that the charge becomes positive. If demand is such that the charge can never become positive, the facility is not justified.[9] The implication for planning for this class of goods is that one should make a forecast of possible usage over the life of the facility. This forecast should be conditional upon the use of the appropriate charges during each period. Given that revenues can cover costs, one selects the appropriate capacity accordingly.[10] Once the facility is constructed, actual prices (as opposed to the forecasted ones) should be simply adjusted by the conventional rule to ration the available supply.

In terms of practical planning of bridges and roads, where it is usually assumed to be not feasible to use direct charges as a rationing device, the implications of the above analysis would seem to indicate that facilities should not be constructed so as to eliminate congestion during all periods. The constraint (8.3) must be binding during some periods in order to have a rational allocation of resources. Since congestion costs serve (imperfectly) as prices, the construction of facilities which never became congested during their useful life should serve as a clear indication of a misallocation of resources.[11] Rational action should not eliminate congestion but merely obtain a (here undefined) appropriate amount of it.

VII. *Concluding Comments*

This paper has not included a discussion of whether the so-called "governmental goods" should really be produced and distributed by the public sector. The issues here are much the same as those discussed for the so-called "public goods." All that is indicated in the previous section is a benchmark analysis against which the behaviors of private, regulated or public actors could be measured. It is true that analytic models of public actors are needed before institutional choices can be made on anything approaching a reasonable basis. Nevertheless, it does appear to be a shame that public goods are called "public."

[9] Note, however, that supply is assumed to be infinitely divisible.

[10] Recall, however, the convenient assumption of no externalities. If there are externalities, then one has to make an adjustment in this rule to take them into account.

[11] One must make an exception here for the very real problem caused by indivisibility in urral areas. Recall the convenient assumption that the size of the facility was continuously divisible.

REFERENCES

1. James M. Buchanan, "Politics, Policy, and the Pigovian Margins," *Economica*, Feb., 1962, pp. 17–28.
2. ———, "An Economic Theory of Clubs," *Economica*, Feb., 1965, pp. 1–14.
3. James M. Buchanan and W. Craig Stubblebine, "Externality," *Economica*, Nov., 1962, pp. 371–84.
4. Ronald H. Coase, "The Federal Communications Commission," *J. of Law and Econ.*, Oct., 1959, pp. 1–40.
5. ———, "The Problem of Social Cost," *J. of Law and Econ.*, Oct., 1960, pp. 1–44.
6. Otto A. Davis, "Economic Elements in Municipal Zoning Decisions," *Land Econ.*, Nov., 1963, pp. 375–86.

7. Otto A. Davis and Andrew B. Whinston, "Some Notes on Equating Private and Social Cost," *S. Econ. J.*, Oct., 1965, pp. 113–26.
8. ———, "Welfare-Economics and the Theory of Second Best," *Rev. of Econ. Studies*, Jan., 1965, pp. 1–14.
9. Harold Demsetz, "The Exchange and Enforcement of Property Rights," *J. of Law and Econ.*, Oct., 1964, pp. 11-26.
10. Jora R. Minasian, "Television Pricing and the Theory of Public Goods," *J. of Law and Econ.*, Oct., 1964, pp. 71–80.
11. Richard A. Musgrave, *The Theory of Public Finance* (McGraw-Hill, 1959).
12. Mancur Olson, *The Logic of Collective Action* (Harvard Univ. Press, 1965).
13. Paul A. Samuelson, "The Pure Theory of Public Expenditure," *Rev. of Econ. and Statis.*, Nov., 1954, pp. 387–89.
14. ———, "Diagrammatic Exposition of a Theory of Public Expenditure," *Rev. of Econ. and Statis.*, Nov., 1955, pp. 350–56.
15. ———, "Aspects of Public Expenditure Theories," *Rev. of Econ. and Statis.*, Nov., 1958, pp. 332–38.
16. ———, "Public Goods and Subscription TV: Correction of the Record," *J. of Law and Econ.*, Oct., 1964, pp. 81–84.
17. Robert H. Strotz, "Two Propositions Related to Public Goods," *Rev. of Econ. and Statis.*, Nov., 1958, pp. 329–31.
18. Richard E. Wagner, "Pressure Groups and Political Entrepreneurs: A Review Article," G. Tullock, ed., *Papers on Non-Market Decision Making* (Charlottesville: Thomas Jefferson Center, Univ. of Virginia, 1966), pp. 161–70.

DISCUSSION

JOHN W. ASHLEY: Professor Alchian has asked me to confine my attention to the paper presented by Professor Nicols; so I shall simply ignore the other papers here.

The Nicols paper is both interesting and confusing. Perhaps the most interesting, and its significant contribution, is the empirical support it lends to the earlier work in this area, particularly that done by Armen Alchian and Reuben Kessel. The most confusing is his method of presentation of his data, particularly the way he relates the data to his theoretical construct. I have nothing but admiration for the tremendous effort he has obviously put forth in the collection and reduction of these data. I hope, therefore, that my comments regarding his method of presentation will not be misinterpreted. Everyone has his own style, and perhaps one should not be faulted on such grounds, but I should like to offer four comments to indicate where his presentation appears to me to be somewhat confusing. I believe the first two are the most important.

1. Professor Nicols' objective is to generate empirical evidence designed to show that monopolies behave differently from competitive enterprises. He therefore develops a restrictive monopoly model, and then asserts that it well represents mutual savings and loan associations (S&L's), which are to be distinguished from stock S&L's, which are considered to be competitive. Unfortunately, Professor Nicols does not rigorously specify the behavior implications of his monopoly and competitive models so that it is never made clear what empirical data are relevant and critical. We are simply told only that "performance is measured in terms of" variables the relevance of which is never made clear.

2. Professor Nicols' argument could also have been clarified considerably if he had specified precisely his null and alternative hypotheses. This probably would have prevented his making possible erroneous implications that because mutual S&L's behave some particular way, stock S&L's are precluded from behaving in exactly the same way. For example, to argue that the managements of mutual S&L's "need not show earnings designed to inflate stock prices" implies that stock S&L's must show such earnings. I do not know whether they do or not, but the point is that such implications should not be simply assumed, and left unspecified and untested.

3. While it is difficult to pin down, I get the impression that Professor Nicols interprets actions by mutual S&L's as stemming almost exclusively from their monopoly power, when it is not at all obvious that other factors are not more important. Monopoly power often may be sufficient cause, but it is not always a necessary cause. It would be helpful to readers if the author would indicate that he had considered the difference between sufficiency and necessity.

4. In analyzing empirical evidence, instead of simply reporting, say, percentage changes in net worth of mutual S&L's compared to stock S&L's, it

would be helpful if the author would discuss what constitutes significant differences in those data. Professor Nicols has not bothered to do this, and I think it detracts from his interpretations. Discussions of tests of significance often are not inappropriate.

It is obvious that Professor Nicols has diligently collected a vast amount of data on a new and important area of research, and he is to be commended for this significant contribution. Too few of us are willing to make the effort required to check up empirically on our theoretical formulations, even though it is such a vital part of the progress toward better understanding of the world in which we live.

PAUL L. KLEINSORGE: My particular assignment is to comment on Mr. Demsetz' paper. Because of the limitations of time, I shall confine my remarks to this task, although the papers submitted by Mr. Nicols, Mr. Davis and Mr. Whinston certainly deserve careful analysis. I feel fortunate in having been assigned Mr. Demsetz' paper, however, because in it he has developed some principles for guiding the construction of an economic theory of property rights, and in my opinion, he has done this very well. I have no serious disagreement with his analysis. My remarks are aimed mainly toward further explanation and clarification.

For instance, in his discussion of the internalization of externalities, Mr. Demsetz uses the example of slaves who offer to buy their freedom by making payment to their masters a sum based upon the expected return to the slaves of being free men. Certainly this would internalize the possible costs and benefits of the situation, but the question arises as to why the masters should ever permit such negotiations in the first place and how the slaves would acquire the wherewithal to make the payment. Slaves cannot buy themselves out, because they have no money with which to complete the exchange. Perhaps we can assume that they can borrow the money in a perfect capital market where there would be individuals willing to lend on the basis of the future income the slaves would earn as free men. They then would become somewhat like indentured servants, and it would be clearer to me that a part of the cost to the master of forced labor is not getting the amount the workers would pay for their freedom—the additional amount above subsistence they could earn in wages if they were free. But Mr. Demsetz is quite right in pointing out that the cost is there. When slaves were freed in the United States, many slave owners considered this act to be a confiscation of their property, and they could assign a figure in dollars to the capital loss, which should correspond to the present value of the future earnings of the freed slaves minus their cost of subsistence

Mr. Demsetz proceeds to discuss the redistribution effects of an exchange of property rights. He states that in a world of zero transaction costs, the output mix that results when the exchange of property rights is allowed is both efficient and independent of who is assigned ownership (except that different wealth distributions may result in different demands). Later he states that if transaction costs are zero, the redistribution of rights will have no allocation effect. Unfortunately, I was unable to gain access to Mr. Coase's article on

"The Problem of Social Cost" to which Mr. Demsetz refers on this point, and perhaps Mr. Demsetz has sufficiently qualified his statement through the exception which he noted. But it seems to me that Mr. Demsetz is in danger of underestimating the income effects in determining the allocation of resources. In the slave example, the freeing of slaves transferred wealth from slave owners to slaves, although in most instances the former slaves still worked for the same employer. It is true that a slave worked for the master for whom his value is greatest, for the master bought the slave from any other master who received less benefit from the slave. But with freedom, what if the alternative for the slave is self-employment? If the value of the slave to the master is (let us say) $3,000 per year, but the slave, if free, would not work voluntarily for less than $5,000, an amount which he feels would compensate him not only for the loss of entrepreneurial income but also for the satisfaction loss of no longer being his own boss, there is an allocative effect. "Who is assigned ownership" and "wealth distribution" seem to go hand in hand. It is difficult to see how they can be separated. Allocation of resources depends upon allocation of income. If "who is assigned ownership" involves allocation, then redistribution of property rights will have an allocative effect.

In his example of the draft, Mr. Demsetz refers to the efficient mix of civilians and military; but there is a unique efficient mix only if there are no income effects and if competition is assumed including a perfect capital market. If Mr. Demsetz' conclusion is that there are no allocative effects with changes in ownership, this conclusion must depend upon an assumption of a perfectly inelastic supply of factors. It ignores the nonmonetary attributes of various jobs, such as the satisfaction of being one's own boss. If such considerations are included, there can be quite different allocations, depending upon the distribution of ownership rights, and each allocation would be efficient for that particular distribution of rights, assuming perfect capital markets, perfect competition, etc. The efficiency, incidentally, would not have to rely on zero transaction costs. It appears to me that it would still be efficient even if transaction costs were above zero.

Beyond this, I am wondering whether Mr. Demsetz has explored the possibility that exchanges of property rights may be motivated, not only by efforts to increase efficiency, but by the concepts of equity which may be involved. In the example that Mr. Demsetz gives of private property rights in land among the American Indians, for instance, was the motivation the desire to increase the output of furs, or was it to prevent certain individuals from appropriating more than their share as they could do if private property rights in land had not developed?

Mr. Demsetz points out that the role of property rights is the internalization of externalities. The externalities comprise possible costs and benefits associated with social interdependencies. It is one of Mr. Demsetz' theses that the emergence of new property rights takes place in response to the desire for adjustment to new benefit-cost possibilities. I believe that Mr. Demsetz has developed this point very well. Since reading his paper, I have begun to expand my own thinking concerning certain situations which I had accepted with little question before. A private person, for instance, is not likely to build

a lighthouse on a reef to warn ships away, because he has no way of collecting payment from the ship owners for this service. The externalities are not internalized. If the private lighthouse owner is allowed by his government, however, to arrange his light so as to entice enemy ships to wreck themselves on the reef, instead of warning them away, he may find that he has quite a lucrative salvage business. The "benefits" are not the same in the two instances, but in the second he has a property right in a business that he could sell. Another example would be a private reforestation project which increases rainfall in a given area, much to the benefit of the owners of farm land nearby. The forest owner in this situation cannot collect from the farmers for the benefit conferred, although the benefit is very real. Since externalities are the benefits and costs of certain people's activities which affect the welfare of others but which are outside of the price system in the market as it is presently structured, a change would have to occur in the situation before property rights in the increased rainfall would appear.

The number of examples appears limitless. Mr. Demsetz' thesis can be applied to all situations involving social interdependencies. It is an approach that has been used too little in the past and should be explored much more in the future. As a tool for analysis, it appears to me that it has much to offer.

HOWARD KUNREUTHER: Davis and Whinston make the important point that the operating characteristics of various proposals for treating public-type goods should be considered before making a decision on which one to adopt. Once these factors are explicitly brought into the analysis, a nonmarket solution may be the most satisfactory one for handling a particular problem.

This result should not seem surprising if our definition of costs is broadened to include the price of information and the expense involved in policing any scheme, as Demsetz has indicated in his paper. In fact, the whole concept of Pareto optimality must then be viewed in a different light from that discussed by Davis and Whinston. It is quite likely that there may be a nonmarket solution which leads to the most efficient allocation of resources, given the technological and institutional constraints of the real world.

By broadly defining these transaction costs to include sociological and political factors, economists are in a position to develop policy recommendations which have a better chance of being implemented than those proposed in the past. For once the costs of eliminating or modifying certain real-world restrictions are made explicit, we may find that it is more efficient to solve the problem within a given institutional framework, just like a firm must combine its factors using some given production function. This is not to say that a theoretical solution based on a frictionless world is unimportant; on the contrary, it provides us with a long-run ideal goal towards which we should strive in formulating "second-best" proposals.

An example of current interest should illustrate this point. Today it is impossible for residents in the U.S. to obtain any private coverage against water damage to their fixed property. They have thus been forced to rely on aid from the Red Cross, federal grants, or low interest loans from the Small Business Administration when their homes or commercial establishments are inun-

dated by either a flood or hurricane. Despite the large number of Corps of Engineer dams and levees that have been erected along the rivers, annual flood losses have been rising over the years. One reason for this increase in damage is that most people are unaware of the potential flood problems when they move into an area, since no insurance is available to explicitly indicate the cost of living there. Congress is thus studying alternative proposals for stemming this tide without relying strictly on federal flood control projects which have the unfortunate side-effect of lulling potential residents into thinking they are safe when often they are not.

How can we, as economists, develop a proposal which will alleviate the flood problem in the future? A logical starting point is to ask the question, "Why hasn't the insurance industry developed some type of protection in the past?" They have been reluctant to do so because the rates in a hazardous area would be so high that homeowners would prefer to take their chances by not purchasing a policy and rely on a sympathetic response from the federal government if they happened to be hit by a disaster.

One possible recommendation would call for the insurance industry to market flood policies at rates reflecting risk while the federal government would declare a hands-off attitude following any future disaster. The homeowner would then have the option of paying the relatively high premium and swimming, or taking his chances and perhaps sinking when the next flood damaged his home. Because the value of his property today is based on a sympathetic federal disaster policy, each resident would have to be given a lump-sum payment to compensate him for the loss in value due to this change in the government's attitude. He would then have the option of selling his house whenever he wanted to at the new market price.

Leaving aside welfare considerations, this proposed remedy is a theoretically optimal one for solving the flood problem if only economic factors are taken into account; however, it is not a feasible one given realistic political and sociological constraints which will be costly to change. The present benevolent attitude toward the less fortunate people in the United States makes it highly unlikely that Congress would vote to discontinue their assistance in the future and force property owners now in the flood plains to pay unusually high rates for insurance protection. Even if such a bill were passed, a family would be reluctant to incur the high readjustment costs of relocating in another area while their house was still intact. They may thus prefer to remain in the flood plain without taking out insurance at high rates, pressuring their congressman for special aid when they suffer losses from the next flood.

Given these restrictions which limit implementation, we should attempt to formulate some politically feasible insurance policy which will be second best in the short run but will still aim in the direction of the theoretically optimal solution. In other words, we want to alleviate the flood problem by decreasing the role of the federal government over time so that residents themselves will eventually pay the price for living in a disaster-prone region.

Without going into detail I would suggest the following basic framework upon which such a system could be based. In order to discourage prospective residents from building in regions where the costs exceed the benefits, premi-

ums on new property should reflect the actual risk of living there. On the other hand, residents who already have homes or businesses in hazardous areas should initially be subsidized on their rates so they could afford to purchase a policy. The federal government would then be in a position to restrict their disaster activities in the future, since insurance would now be available for existing property at a reasonable price. In fact, VA and FHA could even force their mortgagors to take out a flood policy. Since the government would not bail out disaster victims in the future, banks and financial institutions would also want to protect their investment by making flood insurance protection a requirement for a mortgage just as they demand fire and extended coverage today.

If a flood or hurricane causes damage to a home, the owner can rebuild the structure wherever he wants, using payments from the insurance company; in the future, however, he will have to pay the full premium on the amount of damage incurred. For example, if a person suffered $15,000 loss on a $20,000 home, he would pay the actuarial rate on the replaced portion of the dwelling and still be subsidized on the remaining $5,000. Thus, when a family is most willing to relocate their property because it has been damaged, there will be a monetary incentive to move out of the flood plain into a safer area and thus incur a lower insurance cost. Over the long run, all insurance premiums will reflect the true risk of living in an area just as in the theoretically optimal solution.

Assuming the above proposal is a feasible one, there may still be other implementable solutions which also may alleviate the flood problem. In general, the economist faces the problem of how to evaluate institutional limitations when developing policy recommendations. He does not want to become so enmeshed in the bureaucratic structure that his proposals barely change the system. On the other hand, he must be realistic enough to know the environmental constraints under which he has to operate. In this sense, economics must be viewed as an art rather than a science.

TOPICS IN MONEY

SOME IMPLICATIONS OF MONEY SUPPLY ANALYSIS*

By DAVID I. FAND

State University of New York at Buffalo

In money and banking textbooks there is a simple link between bank reserves, deposits, and money. In a world where banks use all their reserves, where there are no free reserves, and where both the banks and the public do not undertake any portfolio changes, there is no need to concern ourselves with the money supply since it is basically a matter of arithmetic. Once we get away from the simple, mechanical link between reserves, deposits, and money, the supply of money has an independent existence as an economic variable determined by behavior and subject to analysis.

A money supply function relates the nominal money supply to a number of policy controlled variables and instruments, to other financial variables (e.g., the bill rate or the rate on time deposits), to nonfinancial variables, and to exogenous variables. This function is derived for a given set of behavior assumptions for the banks and the public and relates changes in monetary base, in reserve requirements, in the discount rate, in other instruments (such as regulation Q) to changes in the nominal supply of money and it may also relate changes in the money supply to changes in other interest rates. More generally, a money supply function may enable us to predict the effects of changes in demand (for money) as well as changes in Federal Reserve actions on the money supply and on interest rates [14].

But while it is desirable to have such a function, there are two kinds of difficulties that need to be considered. One difficulty is statistical in nature. We may not be able to estimate a supply function if the parameters affecting supply also affect demand. In this case the observed data points reflect the joint influence of both and it may not be possible to estimate a supply function. We shall abstract from this problem by assuming that the demand for money is a demand for real balances but that the dependent variable in the supply function is the supply of nominal money balances.[1]

The second kind of difficulty arises from conceptual differences among monetary theorists. Is there any point in estimating a M.S. function if money is a passive variable reflecting primarily movements in the real economy?[2] Is there any point in estimating a function relating M_1 or M_2 to Federal Reserve instruments if other intermediary claims are relatively perfect substitutes for money? Is there any point in estimating the variables determining the money supply if the private spending responds to changes in interest rates and not necessarily to changes in liquidity?

We shall attempt to show that the issues in dispute between those who take a monetary view—that money may be a causal, active, and independent factor in changing the level of output, employment, and the price level [1] [2] [15]—and those who take a real (nonmonetary) view and argue that money is passive changing primarily in response to changes in the real economy, between those who argue that M_1 or M_2 is a well-defined and a reasonably well-behaved quantity [2] [5] [15] and those who would apply the money supply concept to a broader liquidity aggregate [29], between those who argue that policy should focus on a liquidity aggregate [2] [12] [15] and those who argue for focusing on interest rates [6] [23], may be clarified and even partially resolved by defining several money supply concepts with alternative behavior specifications. These concepts suggest a number of M.S. functions both to discriminate between opposing views and to reconcile some of the differences.

In this paper we carry out part of this task. Accordingly, in Section I we define several money supply concepts and spell out the behavior assumptions in each; in Sections II and III we compare these concepts with the M.S. functions that have been estimated in several models, construct additional M.S. functions with alternative specifications, and derive elasticities or multipliers for each concept. We then analyze these elasticities. These elasticities and multipliers are calculated from the structural equations and any defects in these equations will affect our estimates.[3] But, in spite of such limitations, we do believe that the M.S. concepts, the behavior specifications, and the M.S. functions defined and constructed in this paper provide a useful framework for analyzing and interpreting the monetary studies that have been made and also for resolving some of the differences among monetary economists.

---

[2] To save space we shall use symbols for the concepts and the variables that come up repeatedly; see the Glossary in Appendix II.

[3] To avoid misunderstanding we distinguish between the calculated elasticities and multipliers and the estimated structural equations on which they are based. These structural equations are taken from the econometric models analyzed in this paper.

I. *The Four Money Supply Concepts*

To define the concept of a money supply function, we need to spell out the *ceteris paribus* conditions and, in particular, the specific adjustments that we permit the banks and the public. There are, however, alternative ways in which to incorporate portfolio behavior for the banks and the public in a financial model and different investigators adopt different assumptions. Consequently, in order to relate the money supply functions that have been estimated by Brunner-Meltzer (linear theory) [1] [2], DeLeeuw-S.S.R.C. [9], DeLeeuw-condensed [10], DeLeeuw and Turek [27], Goldfeld [16] and Teigen [24] [25] as well as the Brunner-Meltzer nonlinear theory [1] [3] and the reduced form framework used by Cagan [5] and Friedman [15], it is useful to distinguish four money supply concepts. For a given definition of money, say $M_1 = C + D$, these four concepts imply a particular set of assumptions concerning the real economy, the commercial banks, and the public; and they also provide a framework to organize, compare, and interpret the available estimates.

In the table below we sketch the behavior assumptions of the four money supply concepts.

TABLE 1

FOUR MONEY SUPPLY CONCEPTS*

M. S. Concept	Bank Behavior Defined by	Public Behavior Defined by
I. $f(X, r_b; T, C: Y)$†,‡	Free reserve adjustment Supply of demand deposits	Holdings of currency, time deposits and other financial assets are restricted
II. $f(X, r_b, T, C: Y)$ $=g(X, r_b- Y)$ since $T=T(r_b)$ $\quad C=C(r_b)$	Free reserve adjustment Supply of time and demand deposits	Currency and time deposits adjusted in accord with demand functions
III. $f(X, r_b, T, C: Y)$ $=g(X: Y)$ since $r_b=r_b(X)$ $\quad T(r_b)=T(X)$ $\quad C(r_b)=C(X)$	Free reserve adjustment Supply of time and demand deposits Demand for financial assets	Demand deposits and all financial assets are adjusted in accord with demand functions
IV. $f(X, r_b, T, C: Y)$ $=g(X)$ since $Y=Y(X)$	Free reserve adjustment Supply of time and demand deposits Demand for financial assets	All financial assets, nonfinancial assets, all flow may be adjusted in accord with demand functions; supply of assets to banks

* The symbols used in this table are defined in the Glossary in Appendix III.
† This is M. S. I for the (P) specification which requires that T and C are held constant. Other M. S. I specifications are defined in Appendix I.
‡ We could also show explicitly a vector of variables, Z, which are exogenous to both the financial and the real sectors but we assume these variables are incorporated in the functional form.

M.S. I is a short-run supply concept in that it restricts the public's holdings of currency and time deposits but since it does permit free reserves to vary, it gives the supply response under conditions slightly more general than those in the textbook case. M.S. II is a short-run supply concept that brings in additional adjustments; the banks (in some models) can vary the supply of time deposits and the public can adjust their holdings of currency and time deposits. M.S. III is a reduced form equation for the entire financial sector. It may be viewed as a first approximation to a long-run concept which, while constraining the real sector, does bring in bank behavior with respect to their assets as well as their liabilities and portfolio behavior for the public with respect to all financial assets. M.S. IV is a reduced form equation for the entire economy (financial and real). If it is interpreted as a long-run supply concept it should be understood that it indicates the equilibrium stocks of money that will be held for particular sets of policy variables and variables exogenous to both the financial and real sectors.[4]

M.S. $I = f(X, r_b; T, C: Y)$ provides an estimate of the supply response assuming that the banks undertake portfolio adjustments while the public's demand for currency and time deposits is suppressed. While M.S. $II = g(X, r_b: Y)$ is defined as a short-run concept, we do allow some substitution between demand deposits, time deposits, and currency; accordingly, M.S. II provides us with an estimate of the substitution possibilities discussed in [6] [23]. M.S. $III = g(X: Y)$ is a reduced form for the entire financial sector and therefore allows for substitution among the entire range of financial assets while holding the variables in the real economy constant. It provides an estimate of the Gurley-Shaw effect [29]. M.S. $IV = g(X)$ is a reduced form equation for both the financial and real sector and allows for leakages into the real economy.

Those who argue that money is passive, responding primarily to the real economy, would have very little interest in M.S. I, II, and III. On the other hand, those who view money as a causal factor in its own right are obviously concerned with M.S. I, II, and III. Among this group there would nevertheless be considerable disagreement as to the significance of the leakage as we move from M.S. I to M.S. IV. Thus, an analysis of these four concepts has implications bearing on the appropriateness of the active versus passive views of money; and it also has implications for the issues that divide the activists.

These four money supply concepts enable us to structure the money supply functions that have been estimated; but if we were to construct

[4] It is also possible to go one step further and define M.S. V with some of the policy controlled variables or instruments as endogenous variables. Thus Teigen [25] attempts to predict unborrowed high-powered money while Smith [23] relates the discount rate to other interest rates. For a more general approach, see Wood [28].

a framework to take account of all the M.S. constructs that may be calculated from the available econometric models, it would necessitate many more categories.[5] To reduce the number of possibilities to a manageable amount but still emphasize the more important differences concerning the M.S. concepts, the behavior specifications, and the relevant variables in the M.S. functions, we restrict our discussion to $M_1 = C + D$. In Section II we analyze the two short-run money supply functions and in Section III M.S. III and M.S. IV.

II. *Short-Run Money Supply Concepts: M.S. I and M.S. II*

M.S. I is a short-run supply concept. It gives the supply response on the assumption that the public's demand for C and T and other financial assets is restricted. There are several ways to impose *ceteris paribus* conditions on the public's holdings of financial assets and different investigators often define these conditions in a manner most compatible with their model. In addition to these behavior assumptions (for the banks and the public), M.S. I also assumes that all variables in the real sector of the economy, including stocks of real assets and flows such as consumption and investment, are held constant.[6] Elasticities or multipliers calculated for M.S. I sometimes differ because the authors have, either explicitly or implicitly, imposed different *ceteris paribus* conditions which we label the (W), (P), (A), and (CO) specifications. (See Appendix I.)

If M.S. I, which is a function of financial variables, is fairly stable, it would provide some support for the monetary view. For this reason it seems desirable, first, to examine the results of different investigators and, second, to see whether the differences may be reconciled if we make allowance for the alternative behavior specifications.

The Multipliers and Elasticities for M.S. I. In Table 2 we compare the elasticities and multipliers for M.S. I. The four behavior specifications are defined in Appendix I and considering the leakages they permit, DeLeeuw's elasticities or multipliers should be the lowest, Goldfeld's the highest, and Teigen's elasticities or multipliers should be intermediate.[7] For H_u the results conform to our expectations, and the

---

[5] To illustrate, M.S. I has been estimated with alternative restrictions on currency and time deposits. Also, we may estimate these four money supply concepts with the steady-state stock functions or with the short-run stock adjustment functions. In addition, these four M.S. concepts can be defined for $M_1 = C + D$, for $M_2 = M_1 + T$, or for an even broader liquidity total; and for any one definition of money and for any one M.S. concept there are anywhere from 2 to 10 or more elasticities or multipliers that we can calculate. And since we have 8 to 10 investigations to analyze, there are obviously too many permutations to cover in this paper.

[6] It is also possible to estimate an M.S. I function using the stock adjustment equation for free reserves.

[7] The Brunner-Meltzer linear theory has greater leakages in that they assume a more complete adjustment for the banks and permit C and T to vary. This may account for the fact that they have lower estimates, since M.S. I under the (CO) specification approximates M.S. II.

TABLE 2

ELASTICITIES AND MULTIPLIERS FOR M.S. I*

Model	Specification	Elasticities of M_1 with Respect to					Multiplier of M_1 with Respect to				
		H_u	r_a	d	i	r_b	H_u	r_a	d	i	r_b
DeLeeuw (S.S.R.C.).....	(W)	1.00	−.144†	−.314	−.074†	.112	2.98	−7.00	−3.08	−2.26	5.93
DeLeeuw (condensed)...	(W)	1.00	−.078	−.317	−.075	.066	3.00	−3.82	−3.11	−2.28	3.52
DeLeeuw-Turek.........	(W)	1.00	−.086	−.313	−.073	.079	2.99	−4.20	−3.07	−2.24	4.18
Goldfeld..............	(P)	2.28	−.052‡	−.80	−.17	.061‡	7.30	−2.70	−8.31	−5.39	3.25
Teigen 1966..........	(A*)	1.10	−.015§	−.347	−.080	.037§	3.25	−0.71	−3.41	−2.42	1.99
Teigen 1964..........	(A)	1.16	−.230#	−.353	−.084	.250#	3.41	−11.21	−3.46	−2.54	11.21
Brunner-Meltzer.......	(CO)**	.76	.043			.046	2.23	−2.09			2.42
		.84	.004			.012	2.49	−2.19			.656

* The M.S. formulas used to derive these estimates are given in Appendix. All elasticities evaluated at 1962 means.
† The elasticities given in the S.S.R.C. volume of −.348 for r_a and .245 for r_b are incorrect.
‡ Goldfeld reports −.076 for r_a and .222 for r_b. The differences are in part due to his using quarterly data for 1962 [2], and his calculating the elasticity for M.S. II.
§ Teigen 1966 develops a function for D based on R_u. We have developed this into a formula for M based on H_u. He reports −.045 for r_a and .114 for r_b.
# Teigen 1964 reports −.170 for r_a and .195 for r_b but he uses means for the period 1946–59. Teigen also uses the commercial paper rate for the bill rate.
** These estimates are based on (1) P. 271 and (2) P. 49.

same holds for d and t. We also note that the multipliers for r_d are all negative and for r_b all positive. The difference between the largest and smallest elasticity for r_d is approximately .21 but the high estimates are fifteen times as large as the low estimates.

The results for H_u are encouraging since they suggest that the multipliers and elasticities calculated from different models, for different periods from different data (see Appendix II) are in fairly close agreement if we make some allowance for different behavior specifications. The observed differences for the interest rates r_d and r_b, while small absolutely, are nevertheless large relatively and this reflects both differ-

TABLE 3

ELASTICITIES FOR r_d AND r_d FOR M.S. I: CONSTANT SPECIFICATION*

Model	$\epsilon\ (M_1, r_d)$				$\epsilon\ (M_1, r_b)$			
	W	A	P	CO	W	A	P	CO
DeLeeuw†	−.144		−.383		.112		.297	
DeLeeuw condensed.....	−.078		−.206		.066		.174	
DeLeeuw-Turek.........	−.086		−.250		.079		.211	
Teigen 1964†	−.190	−.230	−.468		.202	.250	.509	
Teigen 1966†	−.013	−.015	−.033		.035	.038	.089	
Goldfeld†	−.030		−.056		.025		.061	
Brunner-Meltzer........				−.043				.046
				−.004				.656

* All elasticities evaluated at 1962 means.
† See footnotes for Table 2.

ences due to the structural equations in the model as well as the differences due to the (W), (P), (A), and (CO) behavior specifications.

M.S. I Elasticities: Comparable Behavior Specifications. To separate out model differences from specification differences and to see whether the estimates for r_d and r_b are particularly sensitive to specification, we have constructed Table 3 which shows elasticities on a comparable (W) or (P) specification. We note that (W) elasticities are lowest, that the (P) elasticities are highest and that the (A) elasticities are intermediate. In this sense the order of magnitude observed for H_u, d and t in Table 2 is preserved. Table 3 indicates that the difference between Teigen 1966 and Goldfeld for r_d is reduced considerably when they are constructed in a comparable manner while the difference between Goldfeld and DeLeeuw (corrected) with respect to r_d is not narrowed but increased whether done on a (W) or on a (P) basis.

The comparisons in Table 3 do not succeed in showing that all the differences in interest rate elasticities are due to behavior specification differences. Some of the differences are due to the structural equations in the model which in turn reflect conceptual differences. Thus Teigen's

1966 elasticities calculated from a D function derived from R_u with an (A) specification are not strictly comparable to DeLeeuw's elasticities calculated from an M function based on H_u with a (W) specification. Similarly, Goldfeld's elasticities calculated from an M function derived from H_u with a (P) specification for an M.S. II concept are not comparable to either DeLeeuw's or Teigen's. To compare Goldfeld with De-Leeuw is to compare (W) elasticities with (P) elasticities and to interchange M.S. I and M.S. II.[8] To compare Goldfeld with Teigen is to compare (P) elasticities with (A) elasticities and to interchange an M function based on H_u with a D function based on R_u.[9]

These conceptual differences seem to affect the interest rate elasticities. They do not, however, seem to distort the elasticities for H_u, d and t. With respect to these Federal Reserve instruments, the results are fairly consistent. Since M.S. I is defined so as to permit only short-run leakages, these results would seem to support the monetary view that money is an independent and causal factor.

The M.S. II Elasticities. Suppose we modify the restrictions implied by the behavior specifications (W), (P), (A), or (CO) and we permit the public to adjust their holdings of currency and time deposits and we also permit the banks to determine supply conditions for time deposits. To calculate the money stock response under these conditions we assume: (1) that the banks will adjust their free reserves and, in some models, we introduce the rates paid on time deposits;[10] (2) that the public's actual holdings of currency and time deposits are no longer held constant but can adjust according to their demand functions; and (3) that we may substitute the demand function for T and C to obtain M.S. II.[11] M.S. II is of the form $M = g(X, r_b:Y)$ and, as in the case with M.S. I, we assume that all variables in the real sector and all other financial assets are held constant.

In Table 4 we compare the elasticities for M.S. II. The demand functions for currency and time deposits in these models are not func-

---

[8] In these models the demand for C and T is not functionally related to H_u, d, t, or r_d, and the use of M.S. II elasticities will primarily affect the Treasury bill comparison.

[9] Such comparisons are made in [16] [25].

[10] In some models the quantity of time deposits is determined by an exogenously given r_t; in others by a behavior function for r_t and in Teigen 1964 the quantity of time deposits is not determined.

[11] This is therefore a generalization of M.S. I with specification (P). In principle, one could similarly generalize the (W), (A), and (CO) specifications by substituting the functions for the ratios or derivatives held constant in M.S. I; in practice this can only be done for Teigen who gives an explicit function for the C/D ratio. We view M.S. II as a short-run function. Consequently, since we use the stock equation for free reserves, C and T, we are, in effect, assuming rapid adjustment. It may be inconsistent to suppose that the public will do nothing about their holdings of financial assets but will adjust their C and T holdings rapidly (which carries the suggestion that financial adjustment is a dichotomous process rather than a continuous one). If so, we should use the short-run stock adjustment equations in M.S. I and M.S. II.

TABLE 4

ELASTICITIES FOR M.S. II*

Model	Specification	Elasticities of M_1 with Respect to					
		H_u	r_d	d	t	r_t or r_t max	r_b
DeLeeuw (S.S.R.C.)...	(P)	2.71	−.383	−.833	−.196	.288	.298
DeLeeuw-Turek.......	(P)	2.75	−.252	−.84	−.20	−.42#	.22
Goldfeld†............	(P)	2.28	−.052	−.80	−.17	+.16	.20
Teigen 1966†§........	(A*)	1.10	−.015	−.35	−.080	−.46	−.013
Teigen 1964†‡........	(A)	1.16	−.230	−.353	−.084		.250
Brunner-Meltzer**.....	(CO)	.76	−.043				.046
		.84	−.004				.012

* All elasticities evaluated at 1962 means.
† See footnotes Table 2.
‡ Teigen 1964 M.S. I elasticities are equal to those in M.S. II since T and C/M are held constant.
§ Teigen 1966 elasticities derived from an A* specification assuming that T is fixed while C/D varies.
# In DeLeeuw-Turek model we use r_t max.
** The elasticities are identical with those in Table 2, the (CO) specification being intermediate between the two M.S. concepts.

tions of either H_u, r_d, d or t; and the calculated elasticities for these instruments in M.S. II (for a given specification) do not differ from those for M.S. I. The elasticities with respect to r_b will be changed and we may also calculate elasticities with respect to r_t, the rate on time deposits, and other exogenous variables.[12]

The difference between these elasticities and those for M.S. I, for the variables other than r_b, reflects the fact that the M.S. II calculations in Table 4 are given for the (P) specification. Teigen 1966 does not provide an estimate of the stock demand for T or for C and we calculate the elasticities for M.S. II by using his equation for C/D. Since the M.S. II results for DeLeeuw, Turek, and Goldfeld are on a (P) basis, we compare these with M.S. I calculations on a (P) basis. The changes for the elasticities with respect to r_b are very small. For DeLeeuw it rises from .297 to .298, for Turek from .21 to .22, and for Goldfeld from .061 to .204. For Teigen 1966 it falls from .089 to −.013. (See discussion of Teigen's results in Table 5.)

These results suggest that substitution among demand deposits, time deposits, and currency does not cause significant changes in the elasticities. It may therefore be tempting to conclude that those [6] [13] [23] who concern themselves with the substitution between demand and time deposits are concerned with a minor issue. But since the structural equations for C and T do not contain any of our policy variables or monetary wealth as parameters, this is not entirely surprising. In addition, we note that Brunner-Meltzer, whose (CO) specification permits

[12] Since we are not solving the complete financial model, r_b is still viewed as an exogenous variable and it is not affected by changes in the policy controlled variables.

TABLE 5

ELASTICITIES FOR M.S. III*

Model	Elasticities of M_1 with Respect to					
	H_u	r_d	d	t	r_t or r_t max	r_b
Teigen 1966........	2.22	−.028	−.67	−.15	−.42	
Teigen 1964........	.33	−.066	−.101	−.024		
DeLeeuw-Turek....	.33	−.004	−.102	−.024	−.073	
Goldfeld..........	1.11	−.025	−.386	−.081	−.066	
	ELASTICITY COMPARISONS FOR SEVERAL M.S. CONCEPTS					
DeLeeuw-Turek						
M.S. I............	2.75	−.252	−.84	−.20		.21
M.S. II...........	2.75	−.252	−.84	−.20	−.42	.22
M.S. III..........	.33	−.004	−.10	−.02	−.07	
Teigen 1966						
M.S. I............	1.10	−.015	−.35	−.08		+.037
M.S. II...........	1.10	−.015	−.35	−.08	−.46	−.013
M.S. III..........	2.22	−.028	−.67	−.15	−.42	
M.S. IV...........	3.85	−.051	−1.22	−.28	−.60	
Goldfeld						
M.S. I............	2.28	−.052	−.80	−.17		.06
M.S. II...........	2.28	−.052	−.80	−.17	.16	.20
M.S. III..........	1.11	−.025	−.39	−.08	−.07	
M.S. IV...........	.88	−.020	−.31	−.06	−.18	

* All elasticities evaluated at 1962 means. See footnotes for Table 2.

greater latitude for substitution, do, in fact, have lower elasticities. For these reasons we are inclined to conclude that our results do not reflect so much on the quantitative significance of substitution but rather suggest some needed improvement in the structural equations for currency and time deposits.

III. *Reduced Form M.S. Concepts III and IV*

We also define a money supply concept which measures the movements in the stock of money in response to adjustments in the entire financial sector. To derive this money supply we solve all the equations in the financial sector simultaneously. Variables such as the Treasury bill rate and the rate on time deposits which are endogenous to the financial sector will therefore be determined and can no longer enter as arguments in the money supply function. M.S. III is of the form $M = g(X:Y)$ and measures the supply response due to a change in any of the policy controlled variables including their effect on other endogenous variables in the financial sector.[13]

One major difference between M.S. II and M.S. III in all models is that we now bring in explicitly the demand for demand deposits to de-

[13] More specifically we solve the structural equations to obtain the reduced form and calculate the elasticities or multipliers from the coefficients of the reduced form equations.

termine the equilibrium level of deposits. In addition, in the Turek model, in the DeLeeuw-condensed model, and in the Teigen 1966 model we bring in a term structure equation while in the DeLeeuw-S.S.R.C. model we introduce all seven financial markets.

The M.S. III elasticities are calculated on the assumption that both the banks and the public will make any portfolio adjustment they may choose. These multipliers thus take account of the Gurley-Shaw effect [29] since they allow for substitution among the entire range of financial assets. And by comparing the estimated responses for M.S. III with those in M.S. I or M.S. II we obtain a quantitative estimate of the Gurley-Shaw effect.

In Table 5 we compare elasticities for M.S. III. A change in H_u can affect the money supply directly and through its effect on interest rates such as r_b which is no longer an exogenous variable. Thus if an increase in H_u lowers r_b and raises the demand for C and T, it will reduce the H_u multiplier; and similarly for d, t and r_d. Comparing these elasticities with those for M.S. I and M.S. II, we find that for Goldfeld and Turek but not for Teigen 1966 they tend to decline—sometimes drastically. In Teigen 1966 the C/D behavior is somewhat perverse in that a lowering of say, r_b, lowers the C/D ratio. This accounts in part for the rise in the multiplier.[14]

Finally we define M.S. IV in the form of $M = g(X)$ which measures the movements in the money stock in response to adjustments in both the real and the financial sector. To derive this money supply we must solve all the structural equations in the financial and real sectors simultaneously to obtain the reduced form; and, in consequence, many of the real variables which are exogenous in M.S. III are now determined by the model. This reduced form M.S. gives the equilibrium stocks of money as a function of policy controlled variables and variables which are exogenous to both the financial and real sectors (e.g., family formation).[15] It is, perhaps, the natural if not the only M.S. function to measure for those who view money as a passive variable responding primarily to developments in the real economy.[16]

[14] Teigen 1966 also assumes that C is a relatively inferior good in that the C/D ratio declines with a rise in income. Brunner and Meltzer [1] point out another perverse feature in Teigen's 1964 model is that a rise in the C/M ratio will raise the money supply. This can be corrected by reformulating his M function in terms of H_u, and we have derived the estimates for his 1964 model on this basis.

[15] A variable exogenous to the real sector of one model may be an endogenous variable in another model.

[16] If we think of an M.S. function defined for different kinds of leakages, M.S. IV introduces leakages into the real economy. In this sense it is a natural step from M.S. II and M.S. III. On the other hand, if we think of the M.S. function as reflecting, or incorporating, a behavior response, M.S. IV is, like M.S. III, a hybrid concept and it is perhaps best to think of it as a reduced form function giving equilibrium stocks of M that are associated with different values of the argument. For lack of better terms we may call both the M.S. III and M.S. IV estimates reduced form elasticities.

TABLE 6

ELASTICITIES FOR M.S. IV*

Model	Elasticities of M_1 with Respect to								
	H_u	r_d	d	t	r_t	r_1	r_a	c	r
Teigen 1966.	3.85	−.051	−1.22	−.28	−.60				
Teigen 1964†	.33	−.066	−.101	−.024					
Goldfeld (Stock)....	.88 ·	−.020	−.308	−.065	−.177				
Goldfeld (Short run)‡.....	.43	−.008	−.094	.038					
Cagan§.....	91% 27%							9% 46%	2% 26%
Brunner-Meltzer#..	.647 .347	−.195 −.145				.14 .46	−.034		

* All elasticities evaluated at 1962 means.
† Teigen's 1964 estimates for M.S. IV are identical with those for M.S. III.
‡ Goldfeld's short-run elasticities derived from his impact multipliers.
§ Cagan gives the relative contribution of each determinant to the rate of changes of M_2 for the period 1875–1955. The second set of estimates give the relative contribution to specific cycles in the trend-adjusted rate of change in M_2. See [5, pp. 19, 26] and Appendix II.
\# These estimates are based on [1, p. 280] and [3, p. 59] Brunner and Meltzer use K instead of H_u. See Appendix II.

In Table 6 we present elasticities for M.S. IV, taking account of the interaction between the real and the financial sector. For Teigen 1966 they rise, but this is due to his C/D equation which is inversely related to Y. For Teigen 1964 the M.S. IV results are identical with those for M.S. III because the real sector is not affected by the monetary variables.

For Goldfeld the M.S. IV elasticities are all lower than for M.S. III which is in accord with our expectations since we are permitting greater leakages here. We also include Goldfeld's short-run elasticities in Table 6 and these are consistently below the elasticities derived from his stock equations. (The short-run elasticities are derived from his impact multipliers and provide an estimate of the first period response.) Brunner and Meltzer's elasticities, derived from their nonlinear theory, are not directly comparable to Goldfeld's since they use K instead of H_u and they introduce several interest rates.[17]

Cagan's results are derived from a different framework in which the *ex post* rate of change of the money supply is allocated to three determinants, the rate of change of H, of R/D and of C/M. Cagan esti-

[17] In their nonlinear theory they use

$$K = \log B^a + \sum_{i=1}^{3} \epsilon(m_1, X_i) dX_i / X_i$$

where $B^a = H_u$, $X_1 = d$, $X_2 = t$, $X_3 = C/D$ and m_1 is the monetary multiplier for M_1 and $\epsilon(m_1 X_1)$ is the elasticity of m_1 with respect to X_i.

mates the percentage contribution of a particular determinant in a given period of time and these are not comparable to the elasticity estimates; and he confines his analysis to M_2.

There are at present two few studies available to calculate reliable M.S. IV elasticities. But the available evidence, meager though it may be, does not point to any superiority of M.S. IV over M.S. I, and does not appear to favor a real view over a monetary view. Those who take the view that money is passive, responding primarily to the real economy, have to recognize that this is an assumption rather than a proposition derived from empirical evidence.

Conclusions

In this paper we have defined four money supply concepts with alternative behavior specifications and developed a framework to compare M.S. elasticities calculated from different econometric models. Analysis of Tables 2-6 suggests that while we are far from any consensus there has been a considerable amount of progress. For any one M.S. concept the estimates for H_u, d and t derived from the different models are consistent, while the interest rate elasticities exhibit greater variability, and some instability, as we change either the concept, the model, or the specification. Similarly, as we take one model and compare the estimates for different M.S. concepts, they too behave in a reasonably satisfactory way. With one exception, the estimates decline as well as allow greater leakages. These findings suggest that while there are some obvious defects in some of the structural equations, they may nevertheless be used to construct M.S. functions; and the elasticities calculated from these functions are plausible.

Our analysis points up several areas where additional research effort would be rewarding. The first problem is the estimation of a free-reserve equation for the banks. Following the pioneering work of Meigs [20] and Tobin [26] almost all the models use a free-reserve type equation in constructing their M.S. function. But the free-reserve model that they use is based, implicitly, on the secondary equilibrium analysis developed by Meigs in Chapter 4 of [20] to point up the inadequacies of a free-reserve target as a proximate objective of monetary policy [11]. This model, while adequate for Meigs's purpose, is not the appropriate model for estimating a structural equation, a deposit function, or a money supply function. And indeed when Meigs attempts to estimate a M.S. function he brings in unborrowed reserves explicitly.[18]

---

[18] Professor Milton Friedman has pointed out to me that Meigs presents two models in his book: the secondary equilibrium model of Chapter 4 which he uses to analyze the appropriateness of a free-reserve target for open market operations; and the model in Chapter 5 which (although somewhat implicitly) he uses to estimate a supply function.

In some of the models the free-reserve equation is one of the structural equations estimated, but the form of the equation follows very closely that given by Meigs in Chapter 4. This is appropriate only when the actual change in the free-reserve ratio is zero [7] [8]; otherwise we need to bring in unborrowed reserves. It would also seem desirable to reformulate the free-reserve model so as to develop a deposit supply function directly.[19]

A second problem concerns the specifications of the demand function for currency and time deposits [4] [19]. If we could introduce some of the policy controlled variables or monetary wealth in these two demand functions, we may obtain better estimates of the substitution effects in M.S. II. Also, most of the models estimate these demand functions in nominal terms and do not bring in the price level or changes in the price level. These two functions are of crucial importance in constructing the various M.S. concepts. Any improvements in these functions and especially of the currency equation would obviously improve our M.S. functions and the substitution estimates in M.S. II.[20]

Finally, to test our assumption of rapid adjustment (implicit in our use of stock equations) it would be desirable to estimate stock equations directly. In some of the models the stock equations are derived by solving for the steady-state solution. In doing this we are crucially depending on the coefficient of the lag term which affects all the coefficients in the stock equation. For these reasons it would seem appropriate to undertake a direct estimate of the lags.

There are some other models which could provide the basis for an M.S. function. These include Tobin's work on the theory of assets and portfolio selection [26], related work on bank portfolio behavior by Hester [17] and Pierce [22], Horwich's study of effective reserves [18] and Morrison's study on the liquidity preference of banks [21].

In closing we note that those who deny the usefulness of an M.S. concept presumably rest their case either on the hypothesis that no statistically significant money supply function can be estimated or, if such a function is obtained, it would be related to the M.S. IV concept. Although we do not perform any significance tests, the elasticities and multipliers analyzed here suggest that the M.S. functions based pri-

[19] I am attempting this in a paper which is not yet completed on "The Supply of Deposits in a Free-Reserve Model."

[20] An alternative way to estimate the money supply is to couple a deposit supply function from a more fully articulated free-reserve model with a currency equation—an approach that extends the model in Teigen 1966. The appeal of this method is that it deals explicitly with the two slippages between Federal Reserve action and the money supply. The free-reserve equation deals with the slippage due to bank portfolio behavior while the currency equation deals with the slippage due to the public's portfolio adjustments. The currency equation may also provide a convenient way to pick up the effects that changes in the real economy impart to the money supply in the short run.

marily on financial variables appear to be stable enough to justify further effort toward their refinement and improvement.[21] In particular, we would suggest (1) a testing of the alternative M.S. I specifications, (2) construction of M.S. II functions based on demand functions for C and T which would attempt to bring in policy controlled variables, (3) construction of M.S. III functions to estimate the quantitative effect of substitution and, most importantly, (4) a comparison of these functions with improved estimates of M.S. IV.

APPENDIX I

The Behavior Specifications for M.S. I

The multiplier or elasticity estimates that have been calculated for M.S. I are based on four alternative *ceteris paribus* conditions which we label the W, P, A, and CO specifications. Thus DeLeeuw's elasticity estimates are defined by specification (W) and C/M and T/M are constant.[22] Teigen's estimates are defined by specification (A), that the stock of time deposits and the C/M or C/D ratios are constant. Goldfeld's estimates are defined by specification (P) that the stock of time deposits and currency are constant.[23] The Brunner-Meltzer linear M.S. estimates are defined by specification (CO) that the marginal propensities to hold currency and time deposits are constant. (The marginal propensity relates the change in currency or time deposits to changes in monetary wealth.) In functional form these four specifications may be written as follows:

$(W) \quad M = f(X, r_b; C/M, T/M; Y)$

$(P) \quad M = f(X, r_b; C, T: Y)$

$(A) \quad M = f(X, r_b; C/M, T: Y), \quad \text{or}$

$(A^*) \quad M = f(X, r_b; C/D, T: Y)$

$(CO) \quad M = f(X, r_b, \partial C/\partial M, \partial T/\partial M: Y)$

where X is a vector of policy controlled variables and instruments such as $(H_u, r_d, d, t, \dots)$ and Y is a vector of variables from the real sector of the economy. (See Appendix II for the M.S. formulas.)

These specifications on portfolio behavior are related to the structural equations of the model and often help simplify the construction of the M.S. function. DeLeeuw's free-reserve equation is homogeneous in $(D+T)$, and the behavior specifications in (W) simplify DeLeeuw's M.S. function.[24]

[21] The evidence presented by Brunner and Meltzer for their linear M.S. theory in [2] bears directly on our M.S. I and M.S. II concepts.

[22] DeLeeuw also assumes that Dg/M is constant.

[23] Although Goldfeld seems to describe the concept in the text [16, pp. 190–92], his calculations are based on the behavior assumptions appropriate to M.S. II. Since specification P is most prevalent in standard money and banking texts, we should perhaps label this the standard specification.

[24] Although DeLeeuw's W specification implies that the demand for C and T are proportional to D, he does not appear to use these conditions when he estimates the demand functions for C, T, and D.

Goldfeld's free-reserve equation is not homogeneous in D or T, and it is natural in his model to calculate a money supply function with the specification (P) that the leakages into currency and time deposits are zero as in M.S. I, or to generalize the (P) specification and to take account of these leakages as in M.S. II. In Teigen's 1964 model the money supply is derived from potential money, M^*, and is therefore a function of T and the C/M ratio. In Teigen's 1966 model the money supply is derived from potential deposits, D^*, and is therefore a function of T and the C/D ratio.[25] Teigen's (A) specification is somewhat intermediate between specifications (P) and (W). Finally, Brunner and Meltzer's (CO) specification enables them to estimate the autonomous changes in C and T—two of the variables in their linear M.S. function.[26]

But while the specifications (W), (P), (A), and (CO) may reflect nothing more than the quest for algebraic simplicity, they may also be derived from substantive considerations. Specification (P) extends the standard textbook treatment by bringing in free reserve behavior. Specification (W) goes one step further in bringing in portfolio behavior but limits this behavior by the assumption that C/M, T/M are held constant. Specification (A) holds C/D or C/M and T constant and, while a compromise between the (W) and (P) specifications, may nevertheless be a useful way to deal with the short-run leakages. Finally specification (CO) goes a step further in introducing additional adjustments by permitting the autonomous components of currency and time deposits to vary. These four specifications are thus natural ways in which to widen the scope of the short-run supply function. It is an empirical question to determine which of these four specifications incorporates the most relevant leakages for short-run analysis.

<div align="center">APPENDIX II</div>

<div align="center">*Money Supply Formulas*</div>

1. DeLeeuw (SSRC)

$$M_1 = \frac{H_u}{1 - D/M_1 + .84d(D/M_1 + D_g/M_1) + .82(T/M_1) + (a_0 + a_1 r_b + a_2 r_d)(D/M_1 + T/M_1)}$$

2. DeLeeuw (condensed model)
   Same formula with different coefficients in the free reserve equation

[25] Teigen defines both M^* and D^* in terms of R_u. To compare Teigen's estimates with the others we have expressed his M.S. function in terms of H_u. (See Appendix II.)

[26] Brunner-Meltzer need to estimate the portfolio shifts with respect to currency and time deposits that generate or absorb reserves (independently of monetary expansion or contraction) in order to complete their linear theory. These changes in the monetary base are identified with the autonomous changes in C and T. The (CO) assumptions enable them to estimate the induced changes in C and T and the autonomous changes are obtained as residuals. In their *J.O.F.* article [1] they specify that c_o and t_o, the autonomous changes in C and T, are functions of interest rates. Since they include two of these interest rates in their M.S. functions to estimate the autonomous movements in vault cash, v_o, the regression coefficient for c_o and t_o should be approximately zero. Both Brunner and Meltzer have informed me that they intended to have c_o and t_o as functions of wealth or permanent income.

3. DeLeeuw-Turek
Same formula with different coefficients in the free reserve equation

4. Goldfeld

$$M_1 = \frac{H_u - T(t\gamma_5 + a_1\gamma_7 r_b) - C(1 - d - a_1\gamma_6 r_b) - (a_0 + a_2 r_b + a_3 r_d)}{d + a_1\gamma_6 r_b}$$

5. Teigen 1964

$$M_1 = \frac{(H_u - t\gamma_5 T)(a_0 + a_1 r_b + a_2 r_d)}{d(1 - h - C/M) + C/M}$$

6. Teigen 1966

$$M_1 = \frac{(1 + C/D)(a_0 + a_1 r_b + a_2 r_d)(H_u - t\gamma_5 T)}{d + C/D}$$

7. Brunner and Meltzer
Linear Theory:

$$M_1 = b_0 + b_1(B + L) + b_2 c_0 + b_3 t_0 + b_4 v_0(i)$$

$$\Delta M_1 = b_0 + b_1\Delta(B^a + L + \pi) + b_2\Delta c_0 + b_3\Delta t_0 + b_4\Delta r_b + b_5\Delta r_d$$

Nonlinear Theory:

$$\log M_1 = a_0 + a_1 K + a_2 \log r_d + a_3 \log W/P_a + a_4 \log r_1 + a_5 \log p + a_6 \log Y/\overline{Y}$$

$$\log M_1 = a_0 + a_1 K + a_2 \log r_1 + a_3 \log r_s + a_4 \log r_d + a_5 \log W$$

8. Cagan—Friedman

$$M_2 = \frac{H}{C/M + R/D - C/M \, R/D}$$

$$d \log M_2/dt = d \log H/dt + M/H(1 - R/D)d(-C/M)/dt$$
$$+ M/H(1 - C/M)d(-R/D)/dt$$

Using average values for the 2nd and 3rd terms on the right side, we get $m = h + r + c + e$, where e is the error introduced by the approximation.

APPENDIX II—CONTINUED FROM PAGE 396

DATA USED IN MONEY SUPPLY FORMULAS

Model	Based on the Data for Period	Number of Observations	Financial Sector	Data	Real Sector
S.S.R.C..............	1948 I*–1962 IV	60	Quarterly average of daily or end of month figures		
Condensed..........	1948 I*–1966 IV	60	Quarterly average of daily or end of month figures	S.A.	Annual rates S.A.
Turek..............	1948 I–1964 IV	68	Quarterly average of daily or end of month figures	S.A.	Annual rates S.A.
Goldfield..........	1950 III–1962 II	48	Call date (or approxima-tion to it)	N.S.A.	Quarterly rates N.S.A.
Teigen 1966........	1953 I–1964 IV	48	Call date (or approxima-tion to it)	N.S.A.	Annual rates N.S.A.
Teigen 1964........	1946 IV–1959 IV	49	Call date (or approxima-tion to it)	N.S.A.	Annual rates N.S.A.
B-M (1)...........	1949 I–1962 IV	56	Quarterly data		Annual rates
(2)...........	1949 I–1964 IV	64	Quarterly data	N.S.A.	Annual rates
(3)...........	1919–41, 1952–58	30	Annual data		Annual rates N.S.A.
Cagan.............	1875–1955		Annual and monthly		Annual rates

* Some equations start in 1952 and some start in 1954.

Appendix III

Glossary

M.S. = money supply

$M_1 = C + D$

$M_2 = M_1 + T$

C = currency in circulation outside the Treasury, the Federal Reserve, and the vaults of all commercial banks

D = demand deposits as defined in the "Money Supply" series of the *Federal Reserve Bulletin* or "adjusted" demand deposits for call date data

T = "adjusted" time deposits at all commercial banks

M.S. I = $f(X, r_b, T, C : Y)$

M.S. II = $g(X, r_b : Y)$

M.S. III = $g(X : Y)$

M.S. IV = $g(X)$

X = vector of monetary policy instrument variables; e.g., H_u, d, t, r_d, $r_{t\,max}$

H_u = unborrowed high-powered money = adjusted monetary base $= B_a = R_u + C$

R_u = unborrowed member bank reserves

d = weighted average reserve requirement for demand deposits at member banks subject to reserve requirements

t = weighted average reserve requirement for time deposits subject to reserve requirements

r_d = discount rate

$r_{t\,max}$ = maximum allowable rate on time deposits at commercial banks (regulation Q)

r_b = Treasury bill rate (3 month)

r_c = commercial paper rate

Y = vector of real sector variables (GNP, disposable income, durable consumption, fixed business investment, etc.)

W = DeLeeuw specification: M.S. I = $f(X, r_b, C/M, T/M : Y)$

P = Goldfeld specification: M.S. I = $f(X, r_b, C, T : Y)$

A = Teigen specification: M.S. I = $f(X, r_b, C/D$ or $C/M, T : Y)$

CO = Brunner-Meltzer specification: M.S. I = $f(X, \ r_b, \ \partial C/\partial M, \partial T/\partial M : Y)$

D_g = Treasury deposits at commercial banks

$$M^* = \frac{H_u - t\gamma_5 T}{d(1 - h - C/M) + C/M}$$

$$D^* = \frac{H_u - t\gamma_5 T}{d + C/M}$$

γ_5 = ratio of member bank "adjusted" time deposits to T

$c_o = C - \partial C/\partial M$ = autonomous component of C

$t_o = T - \partial T/\partial M$ = autonomous component of T

$r_t =$ some measure of the interest rate on time deposits in commercial banks (the average rate as estimated by the FDIC in the S.S.R.C., Condensed, and Goldfeld models; a combination of this average up to 1961 and the rate on new CD's after 1961 in the Teigen 1966 model; $r_{t\,max}$ in Turek)

$h =$ ratio of nonmember commercial bank "adjusted" demand deposits to D

$H =$ high-powered money $= R_T + C$

$R_T =$ total member bank reserves

$B + L =$ "extended base" $=$ monetary base plus "liberated reserves"

$v_o =$ autonomous component of vault cash holdings resulting from changes in interest rates

$K =$ defined in text

$W =$ public's nonhuman wealth at current prices

$P_a =$ deflator for W

$r_e =$ interest rate on long-term government bonds

$Y/\overline{Y} =$ index of transitory income

$P =$ income deflator

$\gamma_6 =$ ratio of "adjusted" demand deposits in reserve city banks to D

$\gamma_7 =$ ratio of "adjusted" time deposits in reserve city banks to T

$r_s =$ short-term interest rate

$\pi =$ cumulated sum of vault cash released for absorption in legal reserves

REFERENCES

1. K. Brunner and A. Meltzer, "Some Further Investigations of the Demand and Supply Functions of Money," *J.O.F.*, May, 1964.
2. ———, *An Alternative Approach to the Monetary Mechanism* (G.P.O., 1964).
3. ———, "Liquidity Traps For Money, Bank Credit and Interest Rates" (unpublished paper).
4. P. Cagan, "The Demand for Currency Relative to the Total Money Supply," J.P.E., Aug., 1958.
5. ———, *Determinants and Effects of Changes in the Stock of Money, 1875–1960* (N.B.E.R., 1965).
6. S. Chase and L. Gramley, "Time Deposits in Monetary Analysis," *F.R.B.*, Oct., 1965.
7. R. G. Davis, "Open Market Operations, Interest Rates, and Deposit Growth," *Q.J.E.*, Aug., 1965.
8. ———, "Testing Some Variants of the Free Reserve Hypothesis" (unpublished paper).
9. F. deLeeuw, *A Model of Financial Behavior. The Brookings—Quarterly Econometric Model of the U. S.* (Rand-McNally, 1965).
10. ———, "A Condensed Model of Financial Behavior" (unpublished paper).
11. W. Dewald, "Free Reserves. Total Reserves, and Monetary Control," *J.P.E.*, Apr., 1963.
12. ———, "Money Supply vs. Interest Rates as Proximate Objectives of Monetary Policy," *N.B.R.*, June, 1966.
13. D. I. Fand, "Intermediary Claims and the Adequacy of our Monetary Controls," in *Banking and Monetary Studies* (Irwin, 1963).
14. ———, "A Time Series Analysis of the 'Bills-Only' Theory of Interest Rates," *Rev. of Econ. and Statis.*, Nov., 1966.
15. M. Friedman and A. Schwartz, *A Monetary History of the U. S. 1867–1960* (Princeton, 1963); Appendix B: Proximate Determinants of the Nominal Stock of Money.
16. S. Goldfeld, *Commercial Bank Behavior* (North Holland, 1966).
17. D. Hester, "An Empirical Examination of a Commercial Bank Loan Offer Function," *Yale Econ. Essays*, Spring, 1962.
18. G. Horwich, "Effective Reserves, Credit and Causality in the Banking System of the Thirties," in *Banking and Monetary Studies* (Irwin, 1963).
19. G. Kaufman, *The Demand for Currency, Federal Reserve Staff Economic Studies*.

20. A. J. Meigs, *Free Reserves and the Money Supply* (Chicago, 1962).
21. G. R. Morrison, *Liquidity Preferences of Commercial Banks* (Chicago, 1966).
22. J. L. Pierce, "Commercial Bank Liquidity," *F.R.B.*, Aug., 1966.
23. W. L. Smith, "Time Deposits, Free Reserves, and Monetary Policy" (unpublished manuscript).
24. R. Teigen, "Demand and Supply Functions for Money in the United States: Some Structural Estimates," *Econometrica*, Oct., 1964.
25. ———, "An Aggregated Quarterly Model of the U. S. Monetary Sector" (unpublished paper, 1966).
26. J. Tobin, "Monetary Theory" (unpublished manuscript).
27. J. Turek, "Prediction of Interest Rates Levels, and Financial Stocks with Simulation Techniques" (unpublished paper).
28. J. Wood, *A Model of Federal Reserve Behavior*, *Federal Reserve Staff Economic Studies*.
29. J. G. Gurley and E. S. Shaw, *Money in a Theory of Finance* (Brookings, 1960).

KEYNES AND THE KEYNESIANS:
A SUGGESTED INTERPRETATION

By AXEL LEIJONHUFVUD
University of California, Los Angeles

I

One must be careful in applying the epithet "Keynesian" nowadays. I propose to use it in the broadest possible sense and let "Keynesian economics" be synonymous with the "majority school" macroeconomics which has evolved out of the debates triggered by Keynes's *General Theory (GT)*. Keynesian economics, in this popular sense, is far from being a homogenous doctrine. The common denominator, which lends some justification to the identification of a majority school, is the class of models generally used. The prototype of these models dates back to the famous paper by Hicks [6] the title of which I have taken the liberty of paraphrasing. This standard model appears to me a singularly inadequate vehicle for the interpretation of Keynes's ideas. The juxtaposition of Keynes and the Keynesians in my title is based on this contention.

Within the majority school, at least two major factions live in recently peaceful but nonetheless uneasy coexistence. With more brevity than accurancy, they may be labeled the "Revolutionary Orthodoxy" and the "Neoclassical Resurgence." Both employ the standard model but with different specifications of the various elasticities and adjustment velocities. In its more extreme orthodox form, the model is supplied with wage rigidity, liquidity trap, and a constant capital-output ratio, and manifests a more or less universal "elasticity pessimism," particularly with regard to the interest-elasticities of "real" variables. The orthodoxy tends to slight monetary in favor of fiscal stabilization policies. The neoclassical faction may be sufficiently characterized by negating these statements. As described, the orthodoxy is hardly a very reputable position at the present time. Its influence in the currently most fashionable fields has been steadily diminishing, but it seems to have found a refuge in business cycle theory—and, of course, in the teaching of undergraduate macroeconomics.

The terms of the truce between the two factions comprise two propositions: (1) the model which Keynes called his "general theory" is but a special case of the classical theory, obtained by imposing certain restrictive assumptions on the latter; and (2) the Keynesian special case is nonetheless important because, as it happens, it is more rele-

vant to the real world than the general (equilibrium) theory. To-
gether the two propositions make a compromise acceptable to both
parties, permitting a decent burial of the major issues which almost
everyone has grown tired of debating—namely, the roles of relative
values and of money—and, between them, the role of the interest rate
—in the "Keynesian system." Keynes thought he had made a major
contribution towards a synthesis of the theory of money and "our fun-
damental theory of value" (*GT*, pp. vi-vii). But the truce between the
orthodox and the neoclassicists is based on the common understanding
that his system was *sui generis*—a theory in which neither relative val-
ues nor monetary phenomena are "important."

This compromise defines, as briefly as seems possible, the result of
what Clower aptly calls the "Keynesian Counterrevolution" [4].

II

That a model with wage rigidity as its main distinguishing feature
should become widely accepted as crystallizing the experience of the
unprecedented wage deflation of the Great Depression is one of the
more curious aspects of the development of Keynesianism, comparable
in this regard to the orthodox view that "money is unimportant"—a
conclusion presumably prompted by the worst banking debacle in U.S.
history. The emphasis on the "rigidity" of wages, which one finds in
the New Economics, reveals the judgment that wages did not fall
enough in the early 1930's. Keynes, in contrast, judged that they de-
clined too much by far. It has been noted before that, to Keynes, wage
rigidity was a policy recommendation and not a behavioral assumption
(e.g., [11]).

Keynes's theory was dynamic. His model was static. The method of
trying to analyze dynamic processes with a comparative static appara-
tus Keynes borrowed from Marshall. The crucial difference lies in
Keynes's inversion of the ranking of price- and quantity-adjustment
velocities underlying Marshall's distinction between the "market day"
and the "short run." The initial response to a decline in demand is a
quantity adjustment. Clower's investigation of a system, which re-
sponds to deflationary disturbances in the first instance by quantity
adjustments, shows that the characteristic Keynesian income-con-
strained, or "multiplier," process can be explicated in terms of a gener-
al equilibrium framework [4]. Such a model departs from the tradi-
tional Walrasian full employment model only in one, eminently reason-
able, respect: trading at "false prices"—i.e., prices which do not allow
the realization of all desired transactions—may take place. Transac-
tors who fail to realize their desired sales, e.g., in the labor market,

will curtail their effective demands in other markets. This implies the amplification of the initial disturbance typical of Keynes's multiplier analysis.

The strong assumption of "rigid" wages is not necessary to the explanation of such system behavior. It is sufficient only to give up the equally strong assumption of instantaneous price adjustments. Systems with finite price velocities will show Keynesian multiplier responses to initial changes in the rate of money expenditures. It is not necessary, moreover, to rely on "monopolies," labor unions, minimum wage laws, or other institutional constraints on the utility maximizing behavior of individual transactors in order to explain finite price velocities. Keynes, in contrast to many New Economists, was adamantly opposed to theories which "blamed" depressions on such obstacles to price adjustments. The implied proposition that, if "competition" could only be restored, "automatic forces" would take care of the employment problem was one of his pet hates. Atomistic markets do not mean instantaneous price adjustments. A system of atomistic markets would also show Keynesian adjustment behavior.

In Walrasian general equilibrium theory, all transactors are regarded as price takers. As noted by Arrow, "there is no one left over whose job it is to make a decision on price" [2, p. 43]. The job, in fact, is entrusted to a *deus ex machina:* Walras' auctioneer is assumed to inform all traders of the prices at which all markets are going to clear. This always trustworthy information is supplied at zero cost. Traders never have to wrestle with situations in which demands and supplies do not mesh; all can plan on facing perfectly elastic demand and supply schedules without fear of ever having their trading plans disappointed. All goods are perfectly "liquid," their full market values being at any time instantaneously realizable. Money can be added to such models only by artifice.

Alchian has shown that the emergence of unemployed resources is a predictable consequence of a decline in demand when traders do not have perfect information on what the new market clearing price would be [1, Chap. 31]. The price obtainable for the services of a resource which has become "unemployed" will depend upon the costs expended in searching for the highest bidder. In this sense, the resource is "illiquid." The seller's reservation price will be conditioned by past experiences as well as by observation of the prices at which comparable services are currently traded (*GT*, p. 264). Reservation price will be adjusted gradually as search continues. Meanwhile the resource remains unemployed. To this analysis one need only add that the loss of receipts from its services will constrain the owner's effective demand

for other products—a feedback effect which provides the rationale of the multiplier-analysis of a system of atomistic ("competitive") markets.

To make the transition from Walras' world to Keynes's world, it is thus sufficient to dispense with the assumed tatonnement mechanism. The removal of the auctioneer simply means that the generation of the information needed to coordinate economic activities in a large system where decision making is decentralized will take time and will involve economic costs. No other "classical" assumptions need be relinquished. Apart from the absence of the auctioneer, the system remains as before: (1) individual traders still "maximize utility" (or profit)—one need not assume that they are constrained from bargaining on their own, nor that they are "money illusioned" or otherwise irrational; (2) price incentives are still effective—there is no inconsistency between Keynes's general "elasticity optimism" and his theory of unemployment. When price elasticities are assumed to be generally significant, one admits the potentiality of controlling the activities of individual traders by means of prices so as to coordinate them in an efficient manner. It is not necessary to deny the existence of a vector of nonnegative prices and interest rates consistent with the full utilization of resources. To be a Keynesian, one need only realize the difficulties of finding the market clearing vector.

III

It is a widely held view that the main weaknesses of Keynesian theory derive from Keynes's neglect of the influence of capital and real asset values on behavior (e.g., [8, pp. 9, 11, 17]; [12, p. 636]). It is above all on this crucial point that the standard model has proved to be a most seriously misleading framework for the interpretation of Keynes's theory. This is readily perceived if we compare the "aggregative structures" of the standard model and the *General Theory* model. In either case, we are usually dealing with but three price relations, so that the relevant level of aggregation is that of four-good models:

Standard Model	*General Theory*
Commodities	Consumer goods
Bonds	Nonmoney assets
Money	Money
Labor services	Labor services

The aggregate production function makes the standard model a "one-commodity model." The price of capital goods in terms of consumer goods is fixed. The money wage is "rigid," and the current value of physical assets is tied down within the presumably narrow range of short-run fluctuations in the "real" wage rate. Relative prices are, in-

deed, allowed little play in this construction. "Money" includes only means of payment, while all claims to cash come under the heading of "bonds."

The four-good structure of the *General Theory* is a condensed version of the model of the *Treatise on Money* (*TM*) with its richer menu of short-term assets. All titles to prospective income streams are lumped together in "nonmoney assets." Bond streams and equity streams are treated as perfect substitutes, a simplification which Keynes achieved through some quite mechanical manipulations of risk and liquidity premia (*GT*, Chap. 17). The fundamental property which distinguishes nonmoney assets both from consumables and from money is that the former are "long" while the latter two are "short"— attributes which, in Keynes's usage, were consistently equated with "fixed" (or "illiquid") and "liquid," respectively (cf. *TM*, V:I, p. 248). The typical nonmoney assets are bonds with long term to maturity and titles to physical assets with a very long "duration of use or consumption." Basically, Keynes's method of aggregation differentiates between goods with a relatively high and a relatively low interest elasticity of present value. Thus the two distinctions are questions of degree. As a matter of course, the definition of money includes all types of deposits, since their interest elasticity of present value is zero, but "such instruments as treasury bills" can also be included when convenient (*GT*, p. 167 n.).

Keynes's alleged neglect of capital is attributed to his preoccupation with the short run in which the stock of physical capital is fixed. The critique presumes that Keynes worked with the standard model in which the value of such assets in terms of consumables is a constant. But in Keynes's two-commodity model, this price is, in principle, a short-run variable and, as a consequence, so is the potential command over current consumables which the existing stock of assets represents. The current price of nonmoney assets is determined by expectations with regard to the "stream of annuities" in prospect and by the rate at which these anticipated future receipts are discounted. The relevant rate is always the long rate of interest. In the analysis of short-run "equilibrium," the state of expectation (alias the marginal efficiency of capital) is assumed to be given, and the price of assets then varies with "the" interest rate.

In Keynes's short run, "a decline in the interest rate" and "a rise in the market prices of capital goods, equities, and bonds" are interchangeable descriptions of the same event. Since the representative nonmoney asset is very long-lived, its interest elasticity of present value is quite high. The price elasticity of the output of augmentable income sources is very high. The aggregative structure of this model leaves no

room for elasticity pessimism with regard to the relationship between investment and the (long) rate of interest. It does not even seem to have occurred to Keynes that investment might be exceedingly interest inelastic, as later Keynesians would have it. Instead, he was concerned to convince the reader that it is reasonable to assume that "a moderate change in the prospective yield of capital-assets or in the rate of interest will not involve an indefinitely great change in the rate of investment" (*GT*, p. 252).

The relationship between saving and the interest rate is of less quantitative significance, but Keynes's ideas on the subject are of considerable interest and give some clues to his theory of liquidity preference. The criticisms of his supposed neglect of wealth as a variable influencing behavior have been directed in particular against the *ad hoc* "psychological law" on which he based the consumption-income relation. This line of criticism ignores the "windfall effect" which "should be classified amongst the major factors capable of causing short-period changes in the propensity to consume" (*GT*, pp. 92-94). This second psychological law of consumption states simply that the propensity to consume out of current income will be higher the higher the value of household net worth in terms of consumer goods. A decline in the propensity to consume may, therefore, be caused either by a decline in the marginal efficiency of capital (*GT*, p. 319) or by a rise in the long rate (*GT*, p. 94; *TM*, V:I, pp. 196-97). In the short run the marginal efficiency is taken as given and, so, it is the interest rate which concerns us.

The usual interpretation focuses on the passages in which Keynes argued that "changes in the rate of time-discount" will not significantly influence saving. In my opinion, these well-known passages express the assumption that household preferences exhibit a high degree of intertemporal complementarity, so that the intertemporal substitution effects of interest movements may be ignored. Consequently, the windfall effect of such changes must be interpreted as a wealth effect.

Hicks has shown that the wealth effect of a decline in interest will be positive if the average period of the income-stream anticipated by the representative household exceeds the average period of its planned "standard stream" [7, especially pp. 184-88]. Households who anticipate the receipt of streams which are, roughly speaking, "longer" than their planned consumption streams are made wealthier by a decline in the interest rate. The present value of net worth increases in greater proportion than the present cost of the old consumption plan, and the consumption plan can thus be raised throughout.

This brings our discussion of the *General Theory* into pretty unfamiliar territory. But Keynes's "vision" was of a world in which the in-

dicated conditions generally hold. In this world, currently active households must, directly or indirectly, hold their net worth in the form of titles to streams which run beyond their consumption horizon. The duration of the relevant consumption plan is sadly constrained by the fact that "in the long run, we are all dead." But the great bulk of the "fixed capital of the modern world" is of a very long-term nature (e.g., *TM*, V:II, pp. 98, 364), and is thus destined to survive the generation which now owns it. This is the basis for the wealth effect of changes in asset values.

Keynes's *Gestalt*-conception of the world resembles Cassel's. Cassel used the wealth effect to argue the "necessity of interest" [3], an argument which Keynes paraphrased (*GT*, p. 94). The same conception underlies Keynes's liquidity preference theory of the term structure of interest. Mortal beings cannot hold land, buildings, corporate equities, British consols, or other permanent income sources "to maturity." Induced by the productivity of roundabout processes to invest his savings in such income sources, the representative, risk-averting transactor must suffer "capital uncertainty." Forward markets, therefore, will generally show a "constitutional weakness" on the demand side [7, p. 146]. The relevance of the duration structure of the system's physical capital has been missed by the modern critics of the Keynes-Hicks theory of the term structure of interest rates [10, pp. 14-16] [9, pp. 347-48].

The recent discussion has dealt with the term structure problem as if financial markets existed in a vacuum. But the "real forces of productivity and thrift" should be brought in. The above references to the productivity of roundabout processes (*GT*, Chap. 16) and the wealth effect indicates that they are not totally ignored in Keynes's general theory of liquidity preference. The question why short streams should command a premium over long streams is, after all, not so different from the old question why present goods should command a premium over future goods. Keynes is on classical ground when he argues that the essential problem with which a theory of asset prices must deal derives from the postponement of the option to consume, and that other factors influencing asset prices are subsidiary: "we do not devise a productivity theory of smelly or risky processes as such" (*GT*, p. 215).

IV

Having sketched Keynes's treatment of intertemporal prices and intertemporal choices, we can now consider how "changing views about the future are capable of influencing the quantity of employment" (*GT*, p. vii). This was Keynes's central theme.

"It is by reason of the existence of durable equipment that the eco-

nomic future is linked to the present" (*GT*, p. 146). The price of augmentable nonmoney assets in terms of the wage unit determines the rate of investment. The same price in terms of consumables determines the propensity to consume. This price is the focal point of Keynes's analysis of changes in employment.

If the "right" level of asset prices can be maintained, investment will be maintained and employment at the going money wage stabilized. If a decline in the marginal efficiency of capital occurs, maintenance of the prices of long-lived physical assets and equities requires a corresponding drop in the long rate and thus a rise in bond prices. To Keynes, "the sole intelligible explanation" (*GT*, p. 201) of why this will normally not occur is that bear speculators will shift into savings deposits. If financial intermediaries do not "operate in the opposite direction" (*TM*, V:I, pp. 142-43), bond prices will not rise to the full extent required and demand prices for capital goods and equities will fall. This lag of market rate behind the natural or "neutral" rate (*GT*, p. 243) will be associated with the emergence of excess demand for money—which always spells contraction. "The importance of money essentially flows from its being a link between the present and the future" (*GT*, p. 293).

Contraction ensues because nonmoney asset prices are "wrong." As before, "false prices" reveal an information failure. There are two parts to this information failure: (1) Mechanisms are lacking which would ensure that the entrepreneurial expectations guiding current investment mesh with savers' plans for future consumption: "If saving consisted not merely in abstaining from present consumption but in placing simultaneously a specific order for future consumption, the effect might indeed be quite different" (*GT*, p. 210). (2) There is an alternative "circuit" by which the appropriate information could be transmitted, since savers must demand stores of value in the present. But the financial markets cannot be relied upon to perform the information function without fail. Keynes spent an entire chapter in a mournful diatribe on the Casino-activities of the organized exchanges and on the failure of investors, who are not obliged to hold assets to maturity, to even attempt "forecasting the prospective yield of assets over their whole life" (*GT*, Chap. 12).

Whereas Keynes had an exceedingly broad conception of "liquidity preference," in the Keynesian literature the term has acquired the narrow meaning of "demand for money," and this demand is usually discussed in terms of the choice between means of payment and one of the close substitutes which Keynes included in his own definition of money. Modern monetary theorists have come to take an increasingly dim view of his speculative demand, primarily on the grounds that the underlying assumption of inelastic expectations represents a "special

case" which is unseemly in a model aspiring to the status of a "general theory" [5, pp. 145-51] [13] [8, p. 10] [9, p. 344]. But it is only in the hypothetical world of Walrasian tatonnements that all the information required to coordinate the economic activities of a myriad traders is produced *de novo* on each market day. In any other construction, traders must rely heavily on "memory" rather than fresh information. In the orthodox model, with its interest inelasticity of both saving and investment, there is admittedly no "real" reason why traders' past experiences should be of a narrow normal range of long rates. In Keynes's model, there are reasons. In imperfect information models, inelastic expectations are not confined to the bond market. The explanation of the emergence of unemployed resources in atomistic markets also relies on inelastic expectations. To stress "speculative behavior" of this sort does not mean that one reverts to the old notion of a Walrasian system adjusting slowly because of "frictions." The multiplier feedbacks mean that the system tends to respond to parametric disturbances in a "deviation-amplifying" manner—behavior which cannot be analyzed with the pre-Keynesian apparatus.

A truly vast literature has grown out of the Pigou-effect idea, despite almost universal agreement on its "practical" irrelevance. The original reason for this strange development was dissatisfaction with Keynes's assertion that the only hope from deflation lies "in the effect of the abundance of money in terms of the wage-unit on the rate of interest" (*GT*, p. 253). This was perceived as a denial of the logic of classical theory. Viewing Keynes's position through the glasses of the standard one-commodity model, it was concluded that it could only be explained on the assumption that he had overlooked the direct effect of an increase in real net worth on the demand for commodities (e.g., [11, pp. 269-70] [12, Note K:1]). The one-commodity interpretation entirely misses Keynes's point: that the trouble arises from inappropriately low prices of augmentable nonmoney assets relative to both wages and consumer goods prices. Relative values are wrong. Absolute prices will "rush violently between zero and infinity" (*GT*, pp. 239, 269-70), if price-level movements do not lead to a "correction" of relative prices through either a fall in long rates or an induced rise in the marginal efficiency of capital (*GT*, p. 263). It is hard to see a denial of "our fundamental theory of value" in this argument.

V

We can now come back to the "terms of the truce" between the neoclassicists and the Keynesian orthodox. I have argued that, in Keynes's theory: (1) transactors do maximize utility and profit in the manner assumed in classical analysis, also in making decisions on saving and investment; (2) price incentives are effective and this includes

intertemporal price incentives—changes in interest rates or expected future spot prices (*GT, loc. cit.*) will significantly affect present behavior; (3) the existence of a hypothetical vector of nonnegative prices and interest rates which, if once established, would bring full resource utilization is not denied.

The only thing which Keynes "removed" from the foundations of classical theory was the *deus ex machina*—the auctioneer which is assumed to furnish, without charge, all the information needed to obtain the perfect coordination of the activities of all traders in the present and through the future.

Which, then, is the more "general theory" and which the "special case"? Must one not grant Keynes his claim to having tackled the more general problem?

Walras' model, it has often been noted, was patterned on Newtonian mechanics. On the latter, Norbert Wiener once commented: "Here there emerges a very interesting distinction between the physics of our grandfathers and that of the present day. In nineteenth century physics, it seemed to cost nothing to get information" [14, p. 29]. In context, the statement refers to Maxwell's Demon—not, of course, to Walras' auctioneer. But, *mutatis mutandis,* it would have served admirably as a motto for Keynes's work. It has not been the main theme of Keynesian economics.[1]

---

[1] The paper is an attempt to summarize some of the conclusions of a lengthy manuscript, "On Keynesian Economics and the Economics of Keynes: A Study in Monetary Theory," to be submitted as a doctoral dissertation to Northwestern University.

REFERENCES

GE: John Maynard Keynes, *The General Theory of Employment, Interest and Money* (London, 1936).
TM: ———, *A Treatise on Money,* Vols. I and II (London, 1930).
1. Armen A. Alchian and William R. Allen, *University Economics* (Belmont, Calif., 1964).
2. Kenneth J. Arrow, "Towards a Theory of Price Adjustment," in M. Abramowitz, et. al., *The Allocation of Economic Resources* (Stanford, 1959).
3. Gustav Cassel, *The Nature and Necessity of Interest* (1903).
4. Robert W. Clower, "The Keynesian Counterrevolution: A Theoretical Appraisal," in F. H. Hahn and F. P. R. Brechling, eds., *The Theory of Interest Rates* (London, 1965).
5. William Fellner, *Monetary Policies and Full Employment* (Berkeley, 1946).
6. John R. Hicks, "Mr. Keynes and the 'Classics': A Suggested Interpretation," *Econometrica,* 1937.
7. ———, *Value and Capital,* 2nd ed. (Oxford, 1946).
8. Harry G. Johnson, "The General Theory After Twenty-Five Years," *A.E.R.,* May, 1961.
9. ———, "Monetary Theory and Policy," *A.E.R.,* June, 1962.
10. David Meiselman, *The Term Structure of Interest Rates* (Englewood Cliffs, N.J., 1962).
11. Don Patinkin, "Price Flexibility and Full Employment," *A.E.R.,* 1948, as reprinted in F. A. Lutz and L. M. Mints, eds., *Readings in Monetary Theory* (Homewood, Ill., 1951).
12. ———, *Money, Interest, and Prices,* 2nd ed. (New York, 1965).
13. James Tobin, "Liquidity Preference as Behavior Towards Risk," *Rev. of Econ. Studies,* 1958.
14. Norbert Wiener, *The Human Use of Human Beings,* 2nd ed. (New York, 1964).

UNCERTAINTY AND THE EFFECTIVENESS
OF POLICY*

By WILLIAM BRAINARD
Yale University

Economists concerned with aggregative policy spend a great deal of their time discussing the implications of various structural changes for the effectiveness of economic policy. In recent years, for example, monetary economists have debated at great length whether the rapid growth of nonbank financial intermediaries has lessened the effectiveness of conventional instruments of monetary control. Similarly, in discussions of the desirability of the addition or removal of specific financial regulations the consequences for the effectiveness of policy play an important role. One of the striking features of many of these discussions is the absence of any clear notion of what "effectiveness" is. At times it appears to be simply "bang per buck"—how large a change in some crucial variable (e.g., the long-term bond rate) results from a given change in a policy variable (e.g., open market operation). A natural question to ask is why a halving of effectiveness in this sense should not be met simply by doubling the dose of policy, with equivalent results.

It seems reasonable to suppose that the consequences of a structural change for the effectiveness of policy should be related to how it affects the policy-maker's performance in meeting his objectives. Suppose, for example, that the policy-maker wants to maximize a utility function which depends on the values of "target" variables. If, after some structural change, the policy-maker finds he is able to score higher on his utility function, then presumably the structural change has improved the effectiveness of policy and vice versa. One of the implications of the "theory of policy" in a world of certainty [6] or "certainty equivalence" [1] [3] [4] [5] is that structural changes which simply alter the magnitude of the response to policy do not alter the attainable utility level.[1] Hence such structural changes do not alter effectiveness in the above sense. Another feature of the theory of policy in a world of certainty is that a policy-maker with more instruments than targets is free to discard the excess instruments, and it makes no difference to his performance which ones he discards. These results are crucially dependent on the assumption that the response of target variables to policy in-

[1] Assuming that the levels of instruments do not enter directly into the utility function.

struments is known for certain. Since it is difficult to imagine a real world policy-maker in such an enviable position, it would seem worthwhile to explore the implications of relaxing that assumption. The first two sections of this paper discuss the implications of uncertainty in the response to policy actions for the selection of optimal policy. Optimal policy in the presence of this type of uncertainty is found to differ significantly from optimal policy in a world of certainty. For example, in general all instruments are used, even if there is only one target variable. Analysis of the optimality question also provides some insight into what constitutes effectiveness.

In the third section it is shown that, not surprisingly, the way a structural change alters the effectiveness of policy depends on how it affects both the expected magnitude and the predictability of response to policy actions. The third section goes on to discuss briefly some of the problems involved in assessing the consequences of structural change for the effectiveness of policy when there are several instruments and several targets, and where the structural change affects the response of the system to disturbances as well as policy instruments.

I. One Target—One Instrument

It is instructive to discuss the complications uncertainty creates in a world of one target and one instrument before discussing the problem of optimal use of policy instruments when there are many instruments and targets. Suppose that the policy-maker is concerned with one target variable (y). Assume that y depends linearly on a policy instrument (P)—for example, government expenditures—and various exogenous variables—for example, autonomous investment demand. For our present purposes the impact of exogenous variables may be summarized in a single variable (u).

$$(1) \qquad\qquad y = aP + u$$

where a determines the response of y to policy action.

The policy-maker faces two kinds of uncertainty. First, at the time he must make a policy decision he is uncertain about the impact of the exogenous variables (u) which affect y. This may reflect his inability to forecast perfectly either the value of exogenous variables or the response of y to them. Second, the policy-maker is uncertain about the response of y to any given policy action. He may have an estimate $\bar{a}$ of the expected value of the response coefficient a in (1) above, but he is aware that the actual response of y to policy action may differ substantially from the expected value. At the time of the 1964 tax cut, for example, there was considerable uncertainty over the magnitude of the tax multiplier.

Both types of uncertainty imply that the policy-maker cannot guar-

antee that y will assume its target value (y^*). But they have quite differ-
ent implications for policy action. The first type of uncertainty, if pres-
ent by itself, has nothing to do with the actions of the policy-maker; it
is "in the system" independent of any action he takes. The assumption
that all of the uncertainties are of this type is one of the reasons Theil
and others [1] [3] [4] [5] are able to prescribe "certainty equivalence"
behavior; that is, that the policy-maker should act on the basis of ex-
pected values as if he were certain they would actually occur. Since in
this case the variance and higher moments of the distribution of y do not
depend on the policy action taken, the policy-maker's actions only shift
the location of y's distribution.

In the presence of uncertainty about the response of y to policy ac-
tions, however, the shape as well as the location of the distribution of y
depends on the policy action. In this case the policy-maker should take
into account his influence on the variability of y.[2]

We will assume that the policy-maker chooses policy on the basis of
"expected utility." In particular, we will follow Theil in assuming that
the policy-maker maximizes the expected value of a quadratic utility
function. In the one-target case this is simply:

(2) $U = - (y - y^*)^2$

where y^* is the target value of y.

The assumption of the quadratic enables us to restrict our attention
to the mean and variance of y and to compare our findings directly with
the familiar certainty equivalence results. The assumption of a quadrat-
ic is, of course, subject to the objection that it treats positive and nega-
tive deviations from target as equally important. The use of a fancier
utility function would provide additional reasons for departing from
certainty equivalence.

The precise relationship between policy actions and the variance of y
is not obvious. In (1) above, for example, the policy-maker may believe
that the response coefficient a is a random variable depending on some
unobserved variables, and that it is correlated with u. In that case the y
is a random variable with a variance given by:

(3) $\sigma_y^2 = \sigma_a^2 P^2 + \sigma_u^2 + 2\rho\sigma_a\sigma_u P$

where σ_u^2 and σ_a^2 are the variances of u and a, respectively, and ρ is the
correlation coefficient between u and a.

On the other hand, it is possible to conceive of part of the uncertainty
of y as the consequence of estimation error. Even if the policy-maker

[2] Theil is of course aware that certainty equivalence behavior is not optimal in this case.
In fact, he suggests that the sampling errors in the response coefficients are the most "dan-
gerous" ones for a policy-maker who acts as if all random coefficient matrices coincide with
their expectations [4, p. 74].

regards the "true" population response coefficient a as nonrandom, he may have to base his actions on an estimate of it obtained by fitting equation (1) to sample data. The estimate he uses will be a random variable, and its variance will affect the "variance" of y around its forecast value.[3]

As in the case where the population response coefficient is random, the magnitude of the policy action affects the contribution of this type of uncertainty to the variance of y. In this case, however, the contribution depends on the difference between the policy taken in the forecast period and the average level of policy pursued in the sample period used in estimating a. Assuming the u's are independent over time, a will be uncorrelated with the u for the forecast period and equation (3) may be rewritten:

$$(3') \qquad\qquad \sigma_y^2 = \sigma_a^2(P - \tilde{P})^2 + \sigma_u^2$$

where $\tilde{P}$ is the average P for the sample period on which the estimate of a is based. Although we will use the first formulation to illustrate the significance of uncertainty in the response of y to policy actions, our results can be translated easily for use in the forecast error case.

Assuming the response coefficient is a random variable, we may find the expected utility associated with a given policy action by substituting (1) in (2):

$$(4) \qquad E(U) = - \left[(\bar{y} - y^*)^2 + \sigma_y^2\right]$$

$$= - \left[(\bar{a}P + \bar{u} - y^*)^2 + \sigma_a^2 P^2 + \sigma_u^2 + 2\rho\sigma_a\sigma_u P\right]$$

where $\bar{y}$ and $\bar{u}$ are the expected values of y and u, respectively. There is no reason to suppose that $\bar{u}$ equals zero.

By differentiating (4) with respect to P and setting the derivative equal to zero, the optimal value of P is easily found to be:

$$(5) \qquad\qquad P^* = \frac{\bar{a}(y^* - \bar{u}) - \rho\sigma_a\sigma_u}{\bar{a}^2 + \sigma_a^2}$$

The optimal policy indicated by equation (5) clearly differs from the policy which would be pursued in a world of certainty or of certainty equivalence. The policy-maker should make use of more information than the expected value of the exogenous variables and of the response coefficient of a. Even when a and u are independently distributed, he

---

[3] Hooper and Zellner [2] provide a discussion of the error of forecast for multivariate regression models. In general the variance of the forecast error is $\sigma_F^2 = \sigma^2(xX1^u)[1 + x_F'^{-1}x_F]$ where x is the matrix of sample observations on the independent variables, and x_F is the vector of deviations of the independent variables from their sample means for the forecast period.

should make use of information about the variance of a as well as its mean. If a and u are not independent, he also needs to know their correlation. The assumption of a quadratic utility function does not lead to certainty equivalence except when the policy-maker is certain about the effects of his actions. Another interesting implication of (5) is that it does not in general pay to aim directly at the target. If a and u are independent, for example, equation (5) can be rewritten to show that the fraction of the expected "gap" between $\bar{u}$ and y^* which should be filled by policy action depends only on the coefficient of variation of a:

(5′) $$P^* = g/(1 + V^2)$$

where $g = (y^* - \bar{u})/\bar{a}$, the expected gap, and V equals $\sigma_a/\bar{a}$, the coefficient of variation of a.[4] Only if the policy-maker is absolutely certain about a $(V=0)$ will he close the entire gap; so long as V is finite, he will partially fill the gap.

Some care must be used in interpreting this result. The gap in this context is not the difference between what policy was "last period" and what would be required to make the expected value of y equal to y^*. In the example we have used, the gap is the difference between P equal to zero—the point where the variance of y is least—and the P required to give an expected value of y equal to y^*. If the expected value of a and u and the standard deviation of a remained unchanged for several periods, the optimal policy would also remain unchanged—the policy-maker would not reduce the gap in successive periods. In the case of "forecasting error" the gap is the difference between the average value of policy in the sample used to estimate a and the P which gives $\bar{y}$ equal to y^*. In this case, if the parameter a were reestimated each period and the expected value of u remained the same, policy would be continually revised, making the expected value of y closer and closer to y^* in successive periods.

A natural question to ask is to what extent uncertainty about a affects optimal behavior. Equation (5) indicates that "moderate" uncertainty about a may have a substantial effect. Suppose, for illustrative purposes, that the monetary authority believes that the equation $Y = aM + u$ correctly specifies the relationship between the stock of money (M), an exogenous variable (u), and money income (Y). It puts its staff to work estimating the relationship and obtains an estimate of 5 for the value of a, significant at a "t" level of 2. Further suppose that for the sample of observations on Y and M used to estimate a, the average level of M was \$100. Now suppose that for the next period the desired level of Y is \$650 and the expected value of u is taken to be \$50. In a

---

[4] (5′) can be used for the estimation error case if P^* is interpreted as the deviation of optimal policy from the mean of P during the sample period, and u is defined to include $\bar{a}$ times that mean.

world of certainty M would simply be set at \$600/5 or \$120. Optimal policy in the uncertain situation confronting the policy-maker is most easily found by consulting equation (5'). If M were set at its average value during the sample period the expected value of Y would be \$550, leaving a gap of \$20 (\$100/5).

According to (5'), however, only 80 percent of this gap should be closed with a coefficient of variation of a equal to one-half. Hence it is optimal to set M at \$116 (=\$100+.8×\$20).

When a and u are not independent, the results are slightly more complicated. The policy-maker must now take into account the covariation
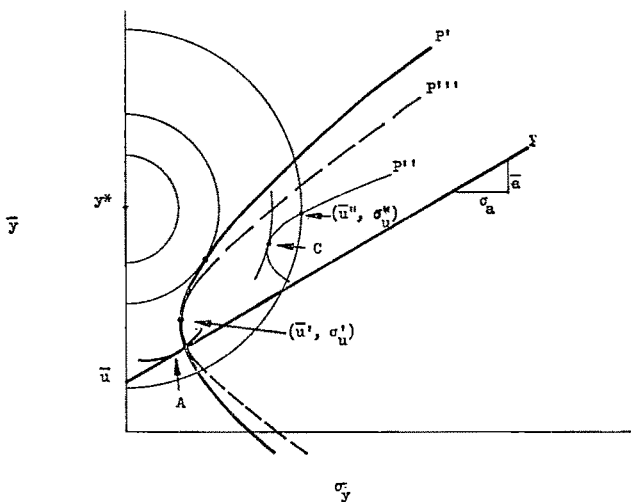
FIGURE 1

between the impact of his policy action and exogenous variables. If there is a positive correlation, it will pay him to shoot for a lower value of $\bar{y}$ than otherwise; if negative, higher. One interesting implication of this is that the fraction of the gap that the policy-maker should close will depend on which side of the target he is on. Perhaps more surprising, if a and u have sufficiently large positive correlation and there is a positive gap or sufficiently large negative correlation and there is a negative gap, it may actually pay for the policy-maker to go the "wrong" way! That is, at the initial point it may actually pay to reduce the variance of y at the expense of increasing the difference between expected y and y^*.

These results can be seen diagrammatically by consulting Figure 1 which shows the expected value of y on the vertical axis and the standard deviation of y on the horizontal axis. Indifference curves, showing various combinations of $\bar{y}$ and σ_y, which have the same expected utility, are drawn "around" y^*, the target value of y. These curves have the

form $(y^*-\bar{y})^2+\sigma_y^2=$ constant; i.e., they are concentric circles drawn around y^*.

In order to focus on the effects of uncertainty concerning the impact of policy, let us assume for the moment that σ_u is zero. Referring to Figure 1, suppose that in the absence of policy action, y is certain to be $\bar{u}$, which is well below y^*. By increasing P by "1 unit," the policy-maker may close the expected gap between y and $\bar{y}$ by an amount $\bar{a}$, but in so doing he also increases the standard deviation of y by σ_a.[5] The line P shows the possibilities open to the policy-maker. Optimal policy corresponds to point A. Because the indifference curves are horizontal where they leave $\bar{y}$ axis, and vertical at a value of $\bar{y}=y^*$, while the slope of the line P is $\bar{a}/\sigma_a$, it always pays to do something, but it never pays to aim for y^*. It is also apparent that reductions in σ_a for a given $\bar{a}$ increase the optimal amount of policy.

These results are not altered for σ_u not equal to zero, so long as the correlation between u and a is 0. In Figure 1 the point $(\bar{u}', \sigma_{u'})$ indicates the expected value and variance of y in the absence of policy action and the line drawn through that point labeled P' indicates the opportunities available to the policy-maker. This "opportunity locus" is curved, reflecting the absence of perfect correlation between a and u. This independence guarantees that unless $\bar{u}=y^*$ it will be optimal to pursue some policy action, for the opportunity locus is vertical at the point $(\bar{u}', \sigma_{u'})$ indicating that the first little bit of policy can be undertaken without increasing the variance of y.[6]

In Figure 1 the line P'' shows the way correlation between a and u alters the opportunity locus available to the policy-maker. In the example shown, a zero level of policy would leave the economy at the point $(\bar{u}'', \sigma_{u''})$. The positive correlation between a and u tilts the locus clockwise through that point so that a small decrease in P will decrease the variance of Y. Optimal policy involves choosing point C, which in this case involves going the "wrong" way. If by chance $\bar{a}(y^*-\bar{u})=\rho\sigma_u\sigma_a$, the policy-maker should do nothing.

Figure 2 shows the way in which the appropriate level of policy action depends on the size of the gap. For given values of $\bar{a}, \rho, \sigma_a, \sigma_u$, the larger the gap, the more the policy-maker should do. With independence of a and u this relationship is linear-homogeneous. When a and u are correlated the relationship is linear, but a zero level of policy is optimal for some nonzero gap.

---

[5] As a convention, the sign of policy is always chosen so that "positive" policy increases $\bar{y}$.

[6] For correlation between u and a equal to 0,

$$\frac{\partial\sigma_y}{\partial P}=\frac{\partial\left[(\sigma_u^2+P^2\sigma_a^2)^{1/2}\right]}{\partial P}=\left.\frac{P\sigma_a^2}{(\sigma_u^2+P^2\sigma_a^2)^{1/2}}\right|_{P=0}=0$$
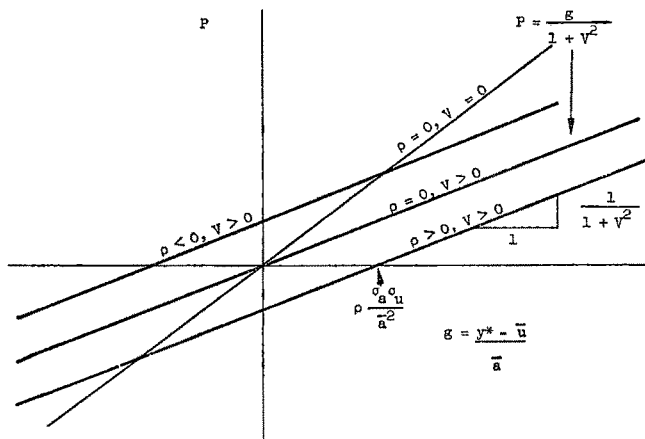
FIGURE 2

II. *Multiple Instruments*

·One of the lessons of the theory of policy under certainty is that, in general, the attainment of n targets requires n instruments. If more than n instruments are available, the policy-maker is free to choose n arbitrarily in order to achieve his objectives. It should already be obvious that this rule breaks down under uncertainty. We will first show that with one target and two instruments it will generally be optimal to use some combination of both instruments.[7] It can similarly be shown that in general all instruments available should be used in pursuing one target. It follows that the addition of an objective requires some sacrifice in performance vis-à-vis objectives already being considered.

Suppose that:

$$(6) \qquad\qquad y = a_1 P_1 + a_2 P_2 + u$$

where again it is desired to maximize the expected value of the utility function given in (2). It will be convenient to define units of policy such that the $\bar{a}_i$ associated with each policy instrument is exactly 1. For simplicity let us assume that correlation between each a_i and u is 0. Then by differentiating the expected value of (2) with respect to P_1 and P_2 we find the conditions for optimal policy to be:

$$(7a) \qquad 0 = (P_1 + P_2 + \bar{u} - y^*) + P_1 \sigma_{a_1}^2 + \rho_{12}\sigma_{a_1}\sigma_{a_2}P_2$$

$$(7b) \qquad 0 = (P_1 + P_2 + \bar{u} - y^*) + P_2 \sigma_{a_2}^2 + \rho_{12}\sigma_{a_1}\sigma_{a_2}P_1$$

where ρ_{12} is the correlation between a_1 and a_2.

---

[7] The idea of looking at the problem of optimal policy behavior in an uncertain world as essentially a problem of portfolio choice arose in discussions with Arthur Okun around 1962.

Solving for the ratio of P_1 to total policy impact (P_1+P_2) we obtain:

$$(8) \qquad P_1/(P_1 + P_2) = \frac{\sigma_{a_2}^2 - \rho_{12}\sigma_{a_1}\sigma_{a_2}}{(\sigma_a^2 - 2\rho_{12}\sigma_{a_1}\sigma_{a_2} + \sigma_{a_1}^2)} \quad {}^8$$

Equation (8) indicates the proportions in which the policy-maker should use the two policy instruments. The optimal policy portfolio as shown in (8) combines the instruments so as to minimize the coefficient of variation of their combined impact. Under the assumption of independence between the a_i's and the u, this policy portfolio can be treated as a single instrument and its optimal level determined as in Section I.

16.a
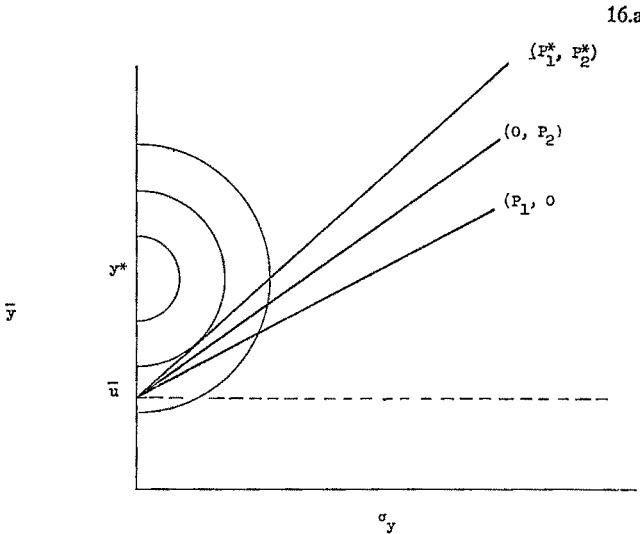
FIGURE 3

The coefficient of variation of the optimal policy package is, of course, less than or equal to the coefficient of variation of any single instrument. Figure 3 shows, for the special case where σ_u and ρ equal zero and where the ratio of the coefficient of variation of a_2 to that of a_1 is .8, the locus for the optimal combination of two policies compared to the loci available for each of the instruments used separately. Figure 3 assumes that the correlation between the impacts of the two instruments is zero. The presence of correlation between the a's complicates the computation of the optimal policy portfolio but does not alter the basic conclusion that several instruments are better than one in the pursuit of one goal.

The optimal amount of P_1 per "unit" of combined policy action

[8] By virtue of the normalization of a_i, of course, the standard deviations are coefficients of variation.

$$\frac{P_1}{P_1 + P_2} = \frac{\sigma_{a_2}}{\sigma_{a_1} + \sigma_{a_2}}$$

$$\frac{P_1}{P_1 + P_2} = \frac{\sigma_{a_2}^2 - \rho \sigma_{a_1} \sigma_{a_2}}{\sigma_{a_2}^2 - 2\rho \sigma_{a_1} \sigma_{a_2} + \sigma_{a_1}^2}$$

$$\rho = \frac{\sigma_{a_2}}{\sigma_{a_1}}$$

$$\sigma_{a_1} > \sigma_{a_2}$$
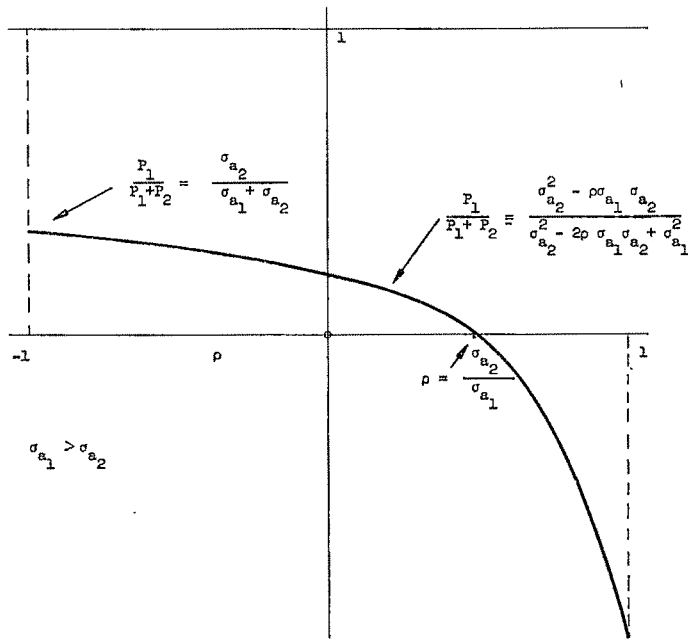
FIGURE 4

$P_1 + P_2$ is shown in Figure 4 as a function of the correlation between a_1 and a_2. In the example shown it is assumed that the coefficient of variation of a_1 is greater than that of a_2. From Figure 4 it can be seen that for sufficiently high positive correlation it is optimal to use P_1 (the "less efficient" instrument) in the "wrong" direction.

Table 1 shows the coefficients of variation for the optimal policy

TABLE 1

COEFFICIENT OF VARIATION FOR ρ_{12} OPTIMAL POLICY PACKAGE

		−9/10	−3/4	−1/2	−1/4	0	1/4	1/2	3/4	9/10	1
	1.0	.22	.35	.50	.61	.71	.79	.87	.94	.97	1.0
	1.1	.23	.37	.52	.64	.74	.83	.90	.97	1.00	0
	1.2	.24	.39	.54	.67	.77	.86	.93	.99	.99	0
	1.6	.27	.43	.61	.74	.85	.93	.99	.98	.85	0
σ_{a_1}	2.0	.30	.47	.65	.79	.89	.97	1.00	.94	.74	0
$\overline{}$	2.4	.31	.49	.69	.82	.92	.99	1.00	.89	.67	0
σ_{a_2}	2.8	.33	.51	.71	.85	.94	.99	.99	.86	.63	0
	3.2	.34	.53	.73	.86	.95	1.00	.98	.83	.60	0
	4.0	.35	.55	.76	.89	.97	1.00	.96	.80	.56	0
	8.0	.39	.60	.81	.93	.99	.99	.92	.73	.49	0

$\sigma_{a_2} = 1.$

package for various values of ρ_{12}, σ_{a_1}, σ_{a_2}. The table is constructed with the variance of a_2 normalized at one and with $\sigma a_2 < \sigma a_1$. Hence the lowest coefficient of variation of policy impact that would be obtained by using a single policy is one. The improvement realized by using both instruments rather than P_2 alone is indicated by the difference between one and the appropriate entry.

The gain from diversification of policy instruments is not a simple function of the correlation coefficient; it also depends on the ratio of coefficients of variation of the two policy instruments. As can be seen from Table 1, negative correlation between a_1 and a_2 greatly assists in reducing the variance in the impact of the policy package, the reduction being greatest for cases where the variances of the two instruments are equal. As the correlation increases, the gain from using two instruments decreases, until at some level of positive correlation none of policy P_1 is used. For correlation greater than that amount, P_1 will be used in the "wrong" direction and some reduction in variance will be realized. In the extreme cases of perfect positive or perfect negative correlation between the a's there exists a policy package with zero variance.

If there is correlation between the disturbance and the impact of either of the policy instruments, it should also be taken into account in the selection of a policy package. Other things being equal, increasing the correlation between the impact of a particular instrument and the disturbance will decrease its use relative to other instruments. The importance of such correlation depends on the size of the gap: the larger the gap, the less the relative contribution of the disturbance to the variance of y after policy action and the less important the correlation of the disturbance and instruments.

The optimal use of n instruments follows the principles illustrated with two; if the error term is independent of the policy response coefficients, the portfolio of instruments which has the lowest coefficient of variation should be chosen. In general this will involve using all of the instruments, and it may involve using some instruments the "wrong" way.

Generalization to many targets is conceptually simple but algebraically tedious. Solution of a multiple-goal problem requires specification of a multidimensional utility function which, if it is quadratic, implicitly provides weights for trading off expected values, variances, and covariances of the policy objectives. The particular solution obviously depends on the weights imbedded in the utility function. One feature of the results is perhaps worth noting: since all policy instruments would be used in pursuit of a single target, improvement in performance vis-à-vis one objective requires sacrificing other objectives—even when the number of instruments exceeds the number of objectives.

III. *Structural Change and the Effectiveness of Policy*

The evaluation of the impact of any particular structural change on the effectiveness of policy is extremely difficult.[9] For our purposes a structural change is described by the way it changes the joint distribution of the parameters a_i and u. The task of determining how the imposition of some new regulation or the emergence of some new financial market alters that distribution is obviously a major one and beyond the scope of this paper. Our relatively modest objective here is to indicate the empirical questions which the above analysis suggests are of importance.

A Single Instrument. First let us consider structural changes which affect only the impact of policy actions; i.e., which do not affect either the expected value or dispersion of the disturbance term. In addition, let us assume that the structural change in question does not alter the correlations among the policy impact coefficients and the disturbance. The effect of such a change on expected utility can be found directly by substituting the value of optimal policy (expressed in terms of $\bar{a}$, σ_a, ρ, $\bar{u}$, etc.) in the equation for expected utility (4). Alternatively, the consequences of structural change can be seen by noting how the "opportunity locus" in Figures 1 and 3 are affected. By substituting $(\bar{y}-\bar{u})/\bar{a}$ for P in equation (3) we find:

$$(9) \qquad \sigma_y = \left(\frac{\sigma_a^2}{\bar{a}^2} (\bar{y} - \bar{u})^2 + 2\rho \frac{\sigma_a}{\bar{a}} \sigma_u(\bar{y} - \bar{u}) + \sigma_u^2 \right)^{1/2}$$

It is immediately clear that a structural change which does not affect the coefficient of variation of a (nor ρ, $\bar{u}$, and σ_u) does not alter the opportunity locus and hence does not alter the expected utility derived from optimal policy. It is also clear that increasing the coefficient of variation, for ρ equal zero, results in an opportunity locus which has a larger variance for every value of $\bar{y}$ (except for zero policy). Hence increasing the coefficient of variation leads to a reduction in the effectiveness of policy. Such a shift is illustrated in Figure 1 by a movement of the opportunity locus labeled P' to the location indicated by the dotted locus labeled P'''.

One of the simplest illustrations of this result is the argument that increasing the legal reserve requirement, thereby reducing the expected response to a given sized change in reserve base, actually increases the effectiveness of monetary control. Suppose one believes that the money

---

[9] What is classified as a structural change obviously depends on the specific problem under discussion. The removal of the ceiling rate on time deposits, for example, would change the response of financial markets to open market operations; that is, it would change the structure within which the Federal Reserve conducts day-to-day policy. At the same time, however, the ceiling rate itself could be used as a policy instrument to influence target variables. While some changes in the "structure" are the direct consequence of actions by the monetary authority, others may be exogenous from their point of view.

stock is all that matters and, further, that the money stock (M) is related to the reserve base (R) by the following equation:

$$(10) \qquad\qquad M = \frac{1}{k + \epsilon} R$$

where k is the legal reserve requirement and ϵ is banks' demand for free reserves (expressed as a fraction of their deposit liabilities). Suppose further that, from the viewpoint of the monetary authority, ϵ is a random variable with a known distribution. Then it is easy to see that, for reasonable distributions of ϵ, increases in k will reduce the expected response of M to R but can reduce the standard deviation of that response more than proportionately, thereby reducing its coefficient of variation. One of the appeals of the 100 percent reserve proposal, of course, is the fact that as k approaches one, the coefficient of variation of the response of M to R goes to zero.

If the correlation between the impact of the policy instrument and u is not zero, matters are slightly more complicated. From (9) it is apparent that increases in σ_a shift the opportunity locus to the right for some values of $(\bar{y} - \bar{u})$ and to the left for others. This is illustrated for a positive ρ in Figure 5 below where the dotted locus corresponds to a higher σ_a. As indicated in Figure 5, the minimum risk that can be obtained is independent of the value of σ_a, and depends only on the value of σ_u and ρ. Hence, it is possible to get the paradoxical result that an increase in the dispersion of response can make policy more "effective" for some set

22a

FIGURE 5
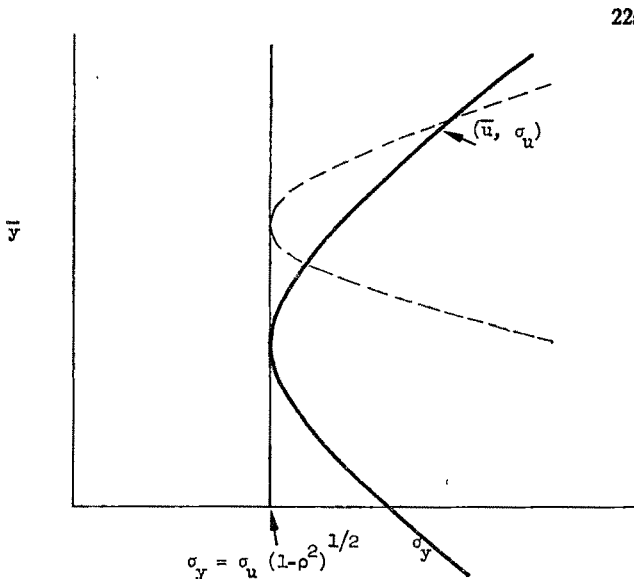
of values of $\bar{u}$ and σ_u. This peculiar result is a consequence of the fact that it is possible to reduce the variance of y below σ_u whenever μ and a are correlated. Consider, for example, the case where a and u are positively correlated, and where $\bar{\mu}$ is less than y^*. In this case a reduction in risk can be obtained for "negative" values of P; i.e., by setting $\bar{y}$ even further from y^* than $\bar{u}$. Suppose that the optimal policy initially involves going in the counterintuitive direction in order to take advantage of such risk reduction. With a larger σ_a the same reduction in risk can be obtained without pushing $\bar{y}$ as far away from y^*; hence it will result in a higher expected utility.

Multiple Instruments. The consequence of structural change when there are a number of instruments is closely analogous to the single instrument case. If the structural change does not affect the coefficient of variation for the individual instruments, their correlations, or the distribution of the disturbance, the opportunity locus is unaltered. If the coefficient of variation of an individual instrument is increased, the consequences can be analyzed in two steps.[10] First, the effect of the change on the coefficient of variation of the optimal policy package can be found. Once this is determined, its consequences for the performance of the policy-maker can be analyzed exactly as in the single instrument case.

The consequences of increasing the coefficient of variation of one instrument are illustrated in Table 1.[11] Two points illustrated in Table 1 are worthy of comment. First, so long as both instruments are used in the same direction (see Figure 4), increasing the coefficient of variation of one instrument increases the coefficient of variation of the optimal policy package. If an instrument is used in the counterintuitive direction, however, increasing its coefficient of variation improves the optimal policy package. Second, the table illustrates the obvious fact that decreasing the effectiveness of one instrument which is being used together with others has much less effect than if the instrument were being used by itself. For example, even with a correlation of $-9/10$, the consequence of increasing σ_{a_1} from 1 to 2 is to increase the coefficient of variation of the optimal policy package by less than 40 percent.

Multiple Targets. The presence of a number of target variables in the utility function greatly complicates the task of evaluating the consequences of a given structural change. It is quite unlikely that a change will affect the opportunity locus for different targets in the same way. Hence the desirability of the change will depend on the relative weights placed on the different targets in the utility function. For example, it is commonly argued that the development of the Euro-dollar market and

[10] Assuming the correlations between the a_i's and u are zero.
[11] Table 1 shows the effect of increasing the coefficient of variation for the less effective of the two instruments.

the general increase in international capital mobility in recent years has made monetary policy a more effective tool in stabilizing the balance of payments, while at the same time reducing its usefulness for controlling domestic economic disturbances.

Disturbances. The discussion thus far has focused on the consequences of structural change for the performance of the policy-maker when confronted with a given distribution of disturbances. A structural change which alters the response to policy actions is also likely to alter the magnitude of the problems with which he must cope. In the above example, greater capital mobility presumably increases the responsiveness of the U. S. balance of payments to disturbances; e.g., in this case to changes in the interest rates in other countries. Thus a structural change which gives the U. S. a more effective tool may also have given it a more difficult task to perform. If this is the case it is not clear that such a structural change is desirable. This can be illustrated by reference to the single instrument case with ρ equal to zero. Suppose for the moment that a structural change doubles $\bar{a}$ and σ_a, but also doubles σ_u, without altering $\bar{u}$. From Figure 1 it is clear that such a change leaves the policy-maker with a worse opportunity locus.[12] If, in addition, there is an increase in the variation of $\bar{u}$ from period to period, still greater losses result.

In the real world, evaluating the desirability of a given structural change is further complicated by the fact that a change which decreases the response of the system to one type of disturbance is likely to increase its response to another. Consider, for example, the consequences of a permanent fixed ceiling rate on time deposits. It can be argued that such a restriction on banks' competition for deposits reduces the response of the system to shifts in the demand of bank borrowers for loans. At the same time, the existence of such a ceiling may increase the destabilizing effect of shifts in depositors' preferences between the liabilities of banks and other financial intermediaries. In general, then, evaluation of some particular structural change requires an empirical judgment as to the relative importance of various kinds of disturbances.

[12] Again, when $\rho \neq 0$, it is possible to have the paradoxical result that increasing σ_u is advantageous.

REFERENCES

1. C. C. Holt, F. Modigliani, J. F. Muth and H. A. Simon, *Planning Production, Inventories, and Work Force* (Prentice-Hall, 1960).
2. John W. Hooper and Arnold Zellner, "The Error of Forecast for Multivariate Regression Models," Cowles Foundation Paper No. 171 (Cowles Found. for Res. in Econ., Yale Univ., 1962).
3. Henri Theil, *Economic Forecasts and Policy*, 2nd ed., Vol. XV of Contributions to Economic Analysis (North-Holland Pub. Co., Amsterdam, 1961).
4. ———, *Optimal Decision Rules for Government and Industry* (Rand McNally, 1964).
5. ———, "Linear Decision Rules for Macrodynamic Policy Problems," *Quantitative Planning of Economic Policy*, Bert G. Hickman, ed. (Brookings Institution, 1965), pp. 18–37.
6. J. Tinbergen, *On the Theory of Economic Policy*, 2nd ed., Vol. I of Contributions to Economic Analysis (North-Holland Pub. Co., Amsterdam, 1952).

DISCUSSION

ALLAN H. MELTZER: I have been requested to comment mainly on David Fand's paper, but I want to make a few remarks about Axel Leijonhufvud's reinterpretation of the *General Theory*. Like many others, I have been puzzled by the differences and apparent inconsistencies between the *Treatise* and the *General Theory* and particularly by the absence from the standard interpretation of the *General Theory* of an effect of changes in relative prices on short-run changes in output. Leijonhufvud's very brief summary of the work that he has done suggests to me that his work is of major importance for macro-theory. I am intrigued particularly by the implications for a theory of dynamics that appear to follow if Walras' auctioneer and the tatonnement process are replaced by an individual who must use resources in his search for market clearing prices and by firms that are "price searchers." The theory outlined in the paper suggests that random shocks, including sudden changes in policy, disturb market clearing prices and the balance between current and future consumption and appears to contain the ingredients for a revised macro-theory of fluctuations that is consistent with neoclassical micro-theory.

Professor Fand asks two main questions in his paper: (1) are the computed responses of the money supply to policy and other variables consistent from one model to another, and (2) given some estimated responses of the nominal stock to the variables on which it depends, how do these responses change when we permit interaction between the money, the nonmoney financial and the real sectors? Fand shows that some of the differences in findings can be explained readily. They are the result of different assumptions about the effect of changes in currency and time deposits on the money supply. When similar assumptions about currency and time deposits are made, some of the differences disappear and the remaining differences can be interpreted as the result of other differences in the models.

What are these other differences? Fand argues—and I agree—that many of them are assumptions about the effects on money resulting from the substitution between money and other financial assets or between money and real assets. Empirical work based on models that incorporate these assumptions provide a basis for testing alternative theories of the monetary process and of the relative effects of policy variables and real variables on money. For example, the empirical estimates based on the various frameworks that Fand considered are capable of discriminating at a first level between three views that I will call (1) the monetary view, (2) the interest rates—"total credit"—financial institutions view, and (3) the "real" or demand view of the supply of money. Let me describe these views.

By the monetary view, I mean that there exist, in principle and in fact, separable, relatively stable demand and supply functions for money and that these functions can be expressed in terms of a few arguments, the proximate determinants of the quantity of money demanded and supplied. It is crucial to the monetary view that a change in a monetary policy variable will induce a

predictable change in the excess supply of money and thus set off a process by which assets are reallocated and/or consumption is changed. Further, this process will continue until the desired stock of real money balances is held; i.e., until the public is willing to hold as real balances the existing nominal stock of money at the new price level.

When the monetary view is stated in this form, there is little room for some of the many disagreements that appear to divide economists. The framework I have described is entirely compatible with the proposition that government deficits financed by interest bearing debt also alter the desired distribution of assets, affect the choice between present and future consumption, the pace of economic activity, and the price level. How, then, does knowledge of the money supply relation contribute to our knowledge as economists?

A partial answer to this question can be obtained by evaluating the statements made by those who deny that there is a meaningful or useful relation between policy variables and the stock of money, using the results in Fand's paper. To make clear that I am not dealing with straw men, I will identify one alternative—the interest rates–total credit–financial institutions view—with Gramley and Chase and another, the "demand" view of the supply of money, with Samuelson.

Gramley and Chase[1] contend that there is no consistent relation between the variables summarizing monetary policy and the stock of money. The size of the money stock, they suggest, depends on the public's demand for time deposits and, in a more general statement of their view (or what appears to be their view) on the demand functions for other close substitutes for money. Although they do not derive these statements from their model[2] and they do not offer a shred of evidence in support of their assertions, they conclude that substitution between monetary assets—whether set off by changes in regulations such as Q or changes in tastes—cause the money supply relation to shift about in an unpredictable way. Comparison of the results for money supply concepts I and II in Fand's paper—whether obtained from the DeLeeuw, Goldfeld, Teigen, or Brunner and Meltzer framework—shows that the Gramley and Chase position is unsupported and at this juncture must be regarded as devoid of empirical content.

This is a finding of some importance. Rarely are we given the opportunity to choose among alternative approaches using evidence as clear cut as the evidence in Fand's paper. By carefully grinding out the implications of the various models, Fand has disposed of the Gramley and Chase argument for the present. The burden of proof is now on Gramley and Chase. As a representative of what they call the "orthodox view," I challenge them to present some valid counterevidence or to abandon their approach.

[1] Lyle Gramley and Samuel Chase, "Time Deposits in Monetary Analysis," *Fed. Res. Bul.*, Oct., 1965, pp. 1380-1406.

[2] Gramley and Chase discuss what they call implications of their framework, although they do not present a formal model or derive any of the so-called "implications." Karl Brunner ("Monetary Analysis and Federal Reserve Policy," multilithed, Ohio State Univ., 1966) has shown that these implications are not implications at all but are assertions that cannot be derived from a hypothesis formulated along the lines they suggest. See, also, William Dewald, "Money Supply Versus Interest Rates as Proximate Objectives of Monetary Policy," *Nat. Bank. Rev.*, June, 1966, pp. 509-22.

Let me make clear that we do not claim—and the evidence does not suggest —that the money supply is unaffected by changes in regulation Q or changes in other variables that affect the desired holdings of time and demand deposits in different ways. As Fand notes, Brunner and I have consistently maintained that interest rates, income, wealth, and other variables affect the ratio of time deposits to demand deposits and that changes in the ratio in turn affect the volume of nominal money supplied. The point at issue is whether the substitutions dominate the effect of policy variables on nominal money and eliminate or substantially reduce the empirical case for a relation connecting policy variable and nominal money stock. This—as I understand it—is the Gramley and Chase position, and it is this position that is rejected by each and every one of the empirical studies of money supply that Fand cites.

Samuelson advises central bankers to ignore the money supply equation.[3] His "policy model" is the standard income-expenditure model, that his textbook made famous, augmented by a demand function for money dependent on income and interest rates. Since the stock of money is a given in this framework, I interpret his model as saying that income is proximately determined by government expenditure and that the stock of money is pulled up and down by changes in income. In this view the stock of nominal money depends on real variables and on interest rates but not on monetary policy variables. Some economists who hold to this view assert that the real variables affect money through the demand for loans. Whatever the process, the stock of money is viewed as independent of policy variables.

Again the evidence fails to support this view. Fand's money supply IV equations incorporate the responses of money to changes in financial and real variables. Once again, there appears to be a persistent effect of policy variables on the stock of money, although the size of the response is generally reduced.

It should be clear that I regard Fand's paper as a useful contribution. It permits the implications just discussed to be drawn from a variety of different frameworks and permits tests of a variety of frequently repeated and often unverified statements. Moreover, it suggests some of the unsolved problems on the supply side that are of importance if we are to improve our ability to predict changes in nominal money balances.

MYRON H. ROSS: After listening to the three papers just presented, I would suggest that the title of this session be changed to "Topics in Uncertainty." These papers deal primarily with the problem of uncertainty and emphasize that it is here to stay. When we shift from the assumption of certainty to uncertainty, our theories may be fundamentally altered. This is the case with the Keynesian model, as Leijonhufvud shows, and this is the emphasis of Brainard's comments on the theory of policy.

Leijonhufvud says it is sufficient to dispense with the auctioneer in order to make the transition from the Walrasian world to the Keynesian world. Once the omniscient and philanthropic auctioneer is removed from the scene, we no

[3] Paul A. Samuelson, "Reflections on Central Banking," *Nat. Bank. Rev.,* Sept., 1963.

longer know what prices will clear markets. The removal of the auctioneer is therefore synonymous with the introduction of uncertainty. Decision-makers are constantly searching for information in a changing environment, with the learning process never being completed.

If information has a positive cost, some information will not be used because its expected marginal cost exceeds its expected marginal revenue. This leads me to question Leijonhufvud's conclusion that other than the removal of the auctioneer, all else in the classical model remains the same: firms can maximize profits, there is no need to assume there are "money illusions," nor do we have to be pessimistic with regard to investment elasticities. My contention is that the removal of the auctioneer means that we remove, of necessity, much more of the classical apparatus.

For example, Leijonhufvud says Keynes's approach leaves no room for elasticity pessimism with regard to investment. I am not sure what "elasticity pessimism" means, but I assume that this probably means that elasticities are no greater than one. Leijonhufvud's statement is based upon the assumption that Keynes only dealt with investment in very durable assets whose present value was sensitive to interest rate changes. This would be true if there were certainty. However, since we are dealing with uncertainty, it seems to me that elasticity optimism is hardly warranted. The further capital asset yields are from the present, the more uncertain they are, so that risk factors tend to swamp the interest rate consideration. Furthermore, uncertainty will usually produce a wide range of estimates of the marginal efficiency of capital by different individuals; and the wider the range of these estimates the lower will be investment elasticity.

This view appears consistent with the *General Theory*. Keynes's consistent repetition of statements like: "it seems unlikely that the influence of banking policy on the rate of interest will be sufficient" (p. 378), or "the economy cannot be made self-adjusting along these lines (of . . . open market operations)" (p. 267) cannot be the basis for the claim that Keynes is referring here to a situation of high elasticities of demand for money only. He is also referring to an inelastic investment function. Even if we assume that the demand for money is highly (but not perfectly) elastic with regard to interest rates, small changes in interest rates should be sufficient to produce enough investment to maintain full employment as long as investment is elastic. Keynes contends this is not so.

Furthermore, Keynes was uncertain about long-term interest rate policy and viewed the matter empirically. He said that interest rates "in normal circumstances may [exercise] a great though not decisive influence on the rate of investment. Only experience, however, can show how management of the rate of interest is capable of continuously stimulating the appropriate volume of investment." Aside from "normal circumstances" (e.g., periods of average uncertainty) one would expect elasticities to be much lower. In any event, econometric studies during the past decade or two appear to have lent support to elasticity pessimism. These studies conclude that investment elasticities usually range from −.3 to −.5.

Leijonhufvud appears to hold that we can rid ourselves of the auctioneer,

without at the same time believing that individuals operate with "money illusions." If you get rid of the auctioneer, however, you will have to accept a certain degree of money illusion. Uncertainty (e.g., removal of the auctioneer) signifies that all facts cannot be taken into account by individuals because information is too costly. Workers, for example, operate with limited information. Money wages are an important piece of information because they are obvious, concrete, and controllable; the cost of living is less obvious and controllable and is thereby deliberately ignored. This means that workers are rational when they offer more labor at higher money wages even though real wages are constant. In terms of the information they have, they behave consistently with their goals. If one contends that workers should get information about the cost of living, one must at the same time argue that the expected marginal revenue from the information will exceed expected marginal costs to the workers.

In the classical model there are no money illusions, witness Pigou's theory of employment. Keynes correctly recognized that there were money illusions for workers. It strikes me that the Keynesian model is deficient in that it is assumed that only workers had a money illusion. Many firms entertain a "profit illusion" when the price level rises and causes depreciation to be understated and profits to be overstated. We should attempt to determine empirically the extent to which these money illusions exist and amend our theory of employment accordingly.

Leijonhufvud says it is possible to remove the auctioneer without at the same time assuming that individuals are constrained from bargaining independently. This also appears to be the contention of Keynes and perhaps indicates that Keynes was not fully aware of how different his model was compared with the classical model. Because the cost of market information is too high, multiprocess and multiperson firms and organizations arise with the function of acting for individuals in order to reduce costs. Large firms and labor unions can reduce costs because there are often economies of scale in the production of information. Thus uncertainty produces large organizations and these organizations have—and usually exercise—market power. This means that we must accept the consequences of large-scale organizations with their sticky prices and long-term contracts. The more sticky prices are the greater is the significance of quantity (or quality) adjustments compared with price adjustments to changes in demand.

Leijonhufvud says one can remove the auctioneer and maintain the assumption that individuals maximize utility or profit. Again, this is consistent with Keynes, though it may only be partially true. When there is uncertainty, profit and utility maximization has an ambiguous quality and it may not be a viable goal. Rather such goals as maximization of sales, minimization of losses, satisfactory profits, or minimization of regret may come to the fore. Money wage rigidity can, for example, be explained via game theory as well as by Keynesian money illusions. One would expect that the greater the uncertainty in the labor market the more conservative workers will be.

The auctioneer, the guarantor of full employment, is like Voltaire's God. If he did not exist, we would invent him. And, indeed Keynes did invent him under the label "fiscal policy" or "the socialization of investment." Keynes did not assume the quantum of uncertainty to be a constant, nor should we.

Though we cannot rid ourselves of uncertainty completely, government policy can significantly reduce uncertainty—or at least not increase uncertainty. In any event, we are a considerable distance from the classical world of theory and policy when we invest our model with the assumption of uncertainty.

Leijonhufvud's demonstration of wealth effects in the *General Theory* is impressive. He shows that a decrease in interest rates will not only cause an increase in investment, but (because net worths increase) also cause an increase in consumption. However, it should be noted that there is in the *General Theory* a secondary wealth effect which will reduce consumption when interest rates fall. A decrease in interest rates increases output and user cost so that net worths will decrease. User cost increases not only because the stock of capital is diminished, but also because the discounted value of future incomes foregone is increased with lower interest rates.

Because I received the papers by Fand and Brainard rather late and have not had sufficient time to do justice to them (though they are quite provocative) and because of the limitations of time, I shall confine myself to a few remarks on each paper. The money supply equations that Fand examines puts the discount rate on a par with the bill rate in determining the money supply. Federal Reserve authorities have frequently contended that banks are extremely reluctant to borrow from the Federal Reserve unless they are temporarily short of reserves. In other words, the discount rate is not an instrument for maximizing profits. However, the elasticities for the discount rate are not very different from those of the bill rate. This fact tends to cast doubt on the "reluctance" doctrine. Still, it would be interesting to compare the level of significance of the regression coefficients of the discount rate and the bill rate.

How useful are the discount rate elasticities for the managers of monetary policy? In the DeLeeuw specification (Table 3) the mean elasticity of the money supply with regard to the discount rate for the six equations is —.09 with a standard deviation of .057. This range of uncertainty is probably too great to permit the practical application of these elasticity values. One can only hope that Fand and others will refine the money supply function so that it will be useful for policy.

Brainard's objection to the "bang per buck" measure of policy effectiveness is logically valid. However, I believe the "bang per buck" approach does have the virtue of being simple and understandable. In answer to his question of why a halving of "effectiveness" should not be met by doubling the dose of policy, it should be indicated that we may not be free to double the dose of policy. There are legal, political, and social constraints on the level of interest rates and the size of the Treasury surplus or deficit.

Brainard demonstrates how the presence of uncertainty makes a fundamental difference in our theory of policy. When the policy-maker is uncertain about the impact of policy on the target, he cannot act as if things were certain. It would be of interest to apply to recent experience the proposition that the policy-maker should design policy so that the coefficient of variation of the combined impact of policies is minimized. Perhaps we can conclude that in 1966 a major policy error was committed when we used monetary policy practically alone with little assistance from fiscal policy, thereby increasing the degree of uncertainty.

EDWARD J. KANE: For want both of time and nimbleness of tongue, I propose to discuss only the Brainard and Fand papers. While they employ different methods and points of view, taken together, these papers are quite complementary, addressing a series of important problems in the theory of monetary control.

Let me begin with Professor Brainard's paper. This is an important piece of work which pioneers in applying the Tobin-Markowitz model of portfolio selection to the choice of an optimal macroeconomic policy mix. It thereby supplies new perspective on long-standing controversies regarding the desirability of various monetary reforms and establishes a new and exciting frontier for the theory of quantitative economic policy.

The new perspective resides in incorporating a second, Friedmanesque dimension to policy effects (the standard deviation of the control error) and in introducing an objective function which allows continuously for trade-offs between the two dimensions of policy effectiveness. This two-dimensionality provides a fresh basis for justifying the refusal of real-world agencies like the Federal Reserve to give up even one of what appears, on Theil-Tinbergen grounds, to be quite redundant instruments. It can also be used to explain why cranks and serious theorists alike have placed continual stress on framework reform: by which I mean proposals to eliminate certain instruments, to sharpen the impact of others, or to tie instrument manipulation to some non-discretionary rule. The key to understanding the persistence of controversy lies in sorting out participants' divergent views of the current shape and location of the relevant P-locus of policy opportunities and, in particular, of the opportunities for bringing about a favorable shift in that locus.

As useful as this perspective should prove to be, it is the territory laid open to future research that I find really exciting. Professor Brainard has pointed up a connection between the theories of control and of stochastic choice. To exploit this link completely, it will be necessary to extend the Brainard model to dynamic situations. For instance, we must explore whether incorporation of Brainard's two-dimensional objective function would reinforce or undermine the Phillips-Baumol work on the vagaries of dynamic economic policy. One would like to know whether taking account of uncertainty would or would not tend to restore the validity of commonsense policy rules. It is hard to imagine that such rules could perform more abysmally than they have in the single-dimensional case.

Let me turn now to Professor Fand's paper, which is also a very promising piece of work. It is related to the preceding article in that it looks ultimately toward determining the structural parameters needed to make the Brainard analysis operational. On the other hand, its immediate goals are much more modest. It is concerned less with estimation (in fact, Professor Fand produces no estimates of his own) than with setting up a classificatory framework whereby we can sort out alternative formulations of the aggregate money-supply function. Also, by converting existing estimates to a common algebraic basis, Professor Fand is able to compare estimates achieved according to quite different formulations.

The paper is weak, however, in two respects: in overall clarity (particularly in its choice of terminology) and in not setting up hard statistical criteria by

which to determine whether alternative estimates are or are not substantially identical. Procedurally, what Professor Fand has done is to subject structural parametric estimates existing in the literature to a series of algebraic transformations. The complexity of some of these transformations explains, no doubt, the absence of statistics measuring goodness of fit. However, without knowing the magnitudes of the relevant standard errors, one cannot possibly know what to make of assertions that estimates are "reasonably close" or "fairly stable." Nor can one feel comfortable with substantive conclusions derived from such presumptions.

That it is often hard to see how Professor Fand got from one set of coefficients to another is due primarily to the poverty of the standard vocabulary by which we can describe differences among closely-related economic models. While this vocabulary problem afflicts the profession as a whole, it has, I think, caused Professor Fand special difficulty.

I object specifically to the way Professor Fand uses four terms: money-supply concepts, differences in specification, model differences, and reduced-form elasticities. I would contend, first, that what Professor Fand labels a money-supply "concept" can be more precisely rendered by the term "mechanism." What is being conceived in this case is a specific process of commercial bank and public reaction to various policy disturbances. Each of Professor Fand's four money-supply concepts features different aspects of some not always fully specified overall pattern of response. Second, what Fand usually means by the phrase "differences in specification" is differences in the algebraic way a particular set of relevant variables is presumed to enter into a given mechanism. In most econometric work, the term specification is used more broadly than this, including (besides differences in algebraic pattern) differences in the set of explanatory variables. Professor Fand calls differences of this latter sort "model differences." Once again, the term he uses seems too inclusive to be precise. Finally, in order to standardize estimated coefficients of reduced-form equations, Professor Fand converts them to elasticities and "for lack of better terms" refers to the resulting quantities as "reduced-form elasticities." I submit that this last choice of words is particularly unfortunate. Ideally, what we want is a term like "implied weighted multiplier" which would indicate simultaneously three things: (1) that the coefficients in question were not estimated directly via a reduced form; (2) that these indirect estimates of reduced-form coefficients have been multiplied by a ratio of sample means; and (3) that they net out all behavioral detail. These coefficients furnish a gross estimate of the total effect of various Federal Reserve actions on the money supply without paying attention to the particular channels through which this effect is presumed to occur.

Differences in algebraic pattern seem not to affect structural estimates for the coefficients of variables under Federal Reserve control, but they do seem to affect structural estimates of interest rate elasticities and estimates of the M.S. III and IV coefficients. This suffices to establish Professor Fand's main point: the money supply is an endogenous variable, but an interesting one. It is neither fully under Federal Reserve control, nor perfectly responsive to developments in the real sector.

AGGREGATE MODELS

A SHORT-RUN AGGREGATE-DEMAND MODEL OF THE INTERDEPENDENCE AND EFFECTS OF MONETARY AND FISCAL POLICIES WITH KEYNESIAN AND CLASSICAL INTEREST ELASTICITIES*

By CARL F. CHRIST

*Johns Hopkins University
and University of Essex*

The purposes of this paper are (1) to emphasize that major domestic monetary and fiscal policies are interdependent because of the existence of the government budget restraint, which allows the authorities only N-1 degrees of freedom in the setting of N monetary and fiscal policy variables; (2) to present a theoretical economic model of aggregate demand that incorporates important features of U.S. domestic monetary and fiscal policy, including the government budget restraint but excluding foreign trade; and (3) to analyze the effects of the four main domestic monetary and fiscal policy variables upon the short-run equilibrium of the economy represented by the model when there is short-run rigidity of the price level.

Results for two of the policy variables—government purchases and the high powered money stock—are as follows. Remember that they are only very approximate, and that they refer to short-run equilibria of real aggregate demand, with rigid prices.

1. The multiplier effect of increased government purchases upon real aggregate demand can be as low as 1.1 (under pure tax finance), or as high as 6.2 (under financing purely by printing high-powered money), or anywhere in between. For example, it is 3.7 when the financing is purely by borrowing from private lenders.

2. The multiplier effect of an increased stock of high-powered money upon real aggregate demand can be as low as 1.7 (where the new money and the resulting induced taxes are spent in buying government bonds from private holders), or as high as 7.9 (where the new money and the resulting induced taxes are spent for increased government purchases of goods and services), or anywhere in between. For example, it is 5.1 if the new money is used purely to replace tax finance, or 2.5 if the new money is used purely to buy government bonds from private holders.

434

This confirms what all good economists know (but most commonly used macro models fail to reflect properly); namely, the effects of fiscal policy depend heavily on how deficit financing is divided between printing money and borrowing from the private sector, and the effects of a change in the high-powered money stock depend heavily on how that change occurs.

The next few paragraphs describe the method of analysis. A ten-equation model of aggregate demand is used. There are two sectors: government (including the treasury and the central bank) and private (including commercial banks and local government).[1] Three assets are explicitly considered: physical capital, government bonds, and high-powered money. Intrasector claims may exist, such as bank deposits and corporate securities, but they cancel out in the consolidated sectors used here.[2]

The ten endogenous variables are c=real consumption, i=real net investment, k=real physical capital, r=yield on perpetual government bonds, r'=yield on physical capital, t=real tax receipts less transfers, w=real private wealth, x=real net national product (NNP), y=real disposable income, and z=real capital gains on government bonds and high-powered money. The four policy variables considered are g=real government purchases, H=high-powered money stock, t_0=autonomous flow of real taxes less transfers, and B=number of government bonds in private hands (each bond being a perpetuity paying \$1.00 per year). The price level P is predetermined and held fixed. So is the marginal tax transfer rate t_1.[3] Stocks are as of the end of the period.

The ten equations are:

(1)	$x = c + i + g$	real NNP definition
(2)	$k = k_{-1} + i$	real physical capital identity
(3)	$w = k + H/P + B/rP$	real private wealth definition
(4)	$t = t_0 + t_1 x$	tax transfer equation
(5)	$y = x - t$	real disposable income definition (excluding capital gain on bonds and high-powered money)

[1] Note that this renders the model incapable of dealing with changes in required reserve ratios, or changes in interest ceilings on bank deposits, unless certain parameters of the model are permitted to vary, because such policies affect the distribution of claims within the private sector and also affect the private sector's aggregate behavior.

[2] The model does exclude any assets (except high-powered money and government bonds) that will arise from claims between the private and government sectors and hence do not cancel out when each sector is consolidated; e.g., Federal Reserve claims against commercial banks as a result of discounting and Treasury deposits in commercial banks. These exclusions are quantitatively minor.

[3] This results in no loss of generality in the analysis, because a change in the marginal tax-transfer rate t_1 can always be linearly approximated by that change in the autonomous tax-transfer rate t_0 which yields the same change in the flow of total taxes less transfers.

(6) $z = H\Delta(1/P) + B\Delta(1/rP)$ real capital gain on bonds and high-powered money

(7) $c = \phi(y, z, w_{-1}, r, r')$ real consumption

(8) $i = \theta(x, k_{-1}, r, r')$ real net investment

(9) $H/P = \psi(x, w, r, r')$ real high-powered money demanded

(10) $B/rP = \beta(x, w, r, r')$ real government bonds demanded

Now government purchases must be financed by some combination of (1) taxes less transfers, (2) printing high-powered money, and (3) borrowing from the private sector, including commercial banks. This requirement is the government budget restraint (11).

(11) $g = t_0 + t_1x + (1/P)\Delta H + (1/rP)\Delta B$

The four policy variables g, t_0, H, and B must be chosen subject to this restraint. Hence only three of them can be exogenous—any three. The fourth one becomes endogenous in a model that incorporates the government budget restraint as its eleventh equation. Thus the model determines short-run equilibrium values of any one policy variable and the other ten endogenous variables, given the values of the other three policy variables, and given also P, t_1, and all the lagged variables. (The equilibrium will not be stationary if the budget is unbalanced, or if net investment is not zero, for the stocks of assets will not then be constant.)

The method of analysis is comparative statics. The first difference of the model is taken, setting $\Delta P = \Delta t_1 = 0$. For simplicity the lagged first differences of all variables are set equal to zero (this does not affect the results very much, for the annual changes in variables rarely exceed 10 percent).

Numerical values are assigned to the initial equilibrium levels of all variables, roughly approximating recent U.S. data, and numerical values are assigned to the partial derivatives of the four behavior functions, based on econometric studies of the U.S. where possible, and otherwise on plausible guesses subject to relevant theoretical restrictions. The first set of values so assigned is labeled Case 1.[4]

[4] They are as follows. Flows in billions of dollars per year: $c = 430$, $i = 50$, $g = t = 120$, $t_0 = 0$, $x = 600$, $y = 480$, $z = 0$. (Note that the budget is assumed to be initially balanced). Stocks in billions of dollars: $H = 60$, $B/r = 240$, $k = 2100$, $w = 2400$. Yields: $r = .0333$ ($3\frac{1}{3}$ percent), $r' = .0667$ ($6\frac{2}{3}$ percent). $P = 1.0$ (index number) and $t_1 = 0.2$ (pure number). Slopes and elasticities:

$$\phi_y = 0.7, \quad \phi_z = 0.07, \quad \eta_{\phi r} = \eta_{\phi r'} = -0.05, \quad \theta_x = 0.1,$$

$$\eta_{\theta r} = -\eta_{\theta r'} = -0.1, \quad \eta_{\psi x} = \eta_{\psi w} = -\eta_{\psi r} = 1, \quad \eta_{\psi r'} = -0.4,$$

$$\eta_{\beta x} = 0.1, \quad \eta_{\beta w} = 1, \quad \eta_{\beta r} = 0.55, \quad \eta_{\beta r'} = -0.25.$$

Sources for these slopes and elasticities are included in the list at the end of the paper. The theoretical restrictions take account of the fact that the demand function for physical capital must be identically equal to total wealth minus the demand functions for real bonds and real high-powered money.

The resulting system of equations for the differences in variables (denoted by Δ's) as between the initial equilibrium and the new one, in Case 1, is as follows:[5]

(1') $\qquad \Delta x = \Delta c + \Delta i + \Delta g$

(2') $\qquad \Delta k = \Delta i$

(3') $\qquad \Delta w = \Delta k + (1/P)\Delta H + (1/rP)\Delta B - (B/r^2 P)\Delta r$

$\qquad\qquad = \Delta k + \Delta H + 30\Delta B - 7200\Delta r$

(4') $\qquad \Delta t = \Delta t_0 + t_1 \Delta x = \Delta t_0 + 0.2\Delta x$

(5') $\qquad \Delta y = \Delta x - \Delta t$

(6') $\qquad \Delta z = -(B/r^2 P)\Delta r = -7200\Delta r$

(7') $\qquad \Delta c = \phi_y \Delta y + \phi_z \Delta z + \phi_r \Delta r + \phi_{r'}\Delta r'$

$\qquad\qquad = 0.7\Delta y + 0.07\Delta z - 645\Delta r - 322.5\Delta r'$

(8') $\qquad \Delta i = \theta_x \Delta x + \theta_r \Delta r + \theta_{r'}\Delta r'$

$\qquad\qquad = 0.1\Delta x - 150\Delta r + 75\Delta r'$

(9') $\quad (1/P)\Delta H = \psi_x \Delta x + \psi_w \Delta w + \psi_r \Delta r + \psi_{r'}\Delta r'$

$\qquad\qquad = \Delta H = 0.1\Delta x + 0.025\Delta w - 1800\Delta r - 360\Delta r'$

(10') $\quad (1/rP)\Delta B = \beta_x \Delta x + \beta_w \Delta w + (\beta_r + B/r^2 P)\Delta r + \beta_{r'}\Delta r'$

$\qquad\qquad = 30\Delta B = 0.04x + 0.1\Delta w + 11160\Delta r - 900\Delta r'$

(11') $\qquad \Delta g = \Delta t_0 + t_1 \Delta x + (1/P)\Delta H + (1/rP)\Delta B$

$\qquad\qquad = \Delta t_0 + 0.2\Delta x + \Delta H + 30\Delta B$

By a policy we shall mean a set of arbitrary values for the changes in any three of the four policy variables, Δg, ΔH, Δt_0, and $30\Delta B$ ($30\Delta B$ is the value of the change in government bonds privately held, since $1/rP = 30$). For any such policy, the foregoing eleven equations will determine the changes in the other eleven variables required to reach the resulting new short-run equilibrium.

Twelve different stimulating policies were chosen for analysis. They are described in the first column of Table 1. For simplicity, the three exogenous changes in policy variables are always chosen to be either 0 or \$1 billion, as shown in three of the columns 2 through 5. The endogenous response of the economy to each of the twelve policies, under the conditions of Case 1, is shown in the other columns, including the endogenous response of the remaining policy variable.

For example, consider policy c. It consists of a \$1 billion open market purchase of government bonds ($30\Delta B = -1$ in column 4), with no

[5] If real output were regarded as fixed by the full employment of given resources, and the price level were regarded as endogenous, then one would set $\Delta x = 0$ rather than $\Delta P = 0$, and these equations would contain ΔP rather than Δx.

TABLE 1

EFFECTS OF TWELVE SELECTED MONETARY AND FISCAL POLICIES UPON THE SHORT-RUN EQUILIBRIUM
VALUES OF THE ELEVEN DEPENDENT VARIABLES IN CASE 1

| Policy | Changes in Policy Variables | | | |
	Δt_0	Δg	$30\Delta B$	ΔH
(1)	(2)	(3)	(4)	(5)
(a) a switch from autonomous tax finance to a combination of induced tax finance and printing high-powered money..	−1	0	0	.50
(b) an increase in government purchases, financed by printing high-powered money except to the extent that induced taxes less transfers contribute to the financing.	0	1	0	.39
(c) an open market purchase of government bonds, financed by printing high-powered money except to the extent that induced taxes less transfers contribute to the financing...	0	0	−1	.75
(d) a switch from autonomous tax finance to a combination of induced tax finance and borrowing from the private sector..	−1	0	.66	0
(e) an increase in government purchases, financed by borrowing from the private sector except to the extent that induced taxes less transfers contribute to the financing...	0	1	.52	0
(f) printing high-powered money, and using it and the induced taxes less transfers to buy government bonds from the private sector; it is the same as policy (c) except for a scale factor..	0	0	−1.33	1
(g) an increase in government purchases, financed entirely by autonomous and induced taxes less transfers (the balanced budget multiplier case).............	.78	1	0	0
(h) a switch from tax finance to borrowing from the private sector; it is the same as policy d except for a scale factor..	−1.51	0	1	0
(i) a switch from tax finance to printing high-powered money; it is the same as policy a except for a scale factor..	−2.01	0	0	1
(j) a switch from debt finance to printing high-powered money, with no change in total taxes less transfers (any induced increase being offset by an equal autonomous cut)...	−.50	0	−1	1
(k) an increase in government purchases, financed entirely by printing high-powered money, with no change in total taxes less transfers (an induced increase being offset by an equal autonomous cut)..	−1.23	1	0	1
(l) an increase in government purchases, financed entirely by borrowing from the private sector, with no change in total taxes less transfers (any induced increase being offset by an equal autonomous cut)...............................	−.73	1	1	0

NOTES: Numerical results were obtained by solving equations (1′)-(11′) for each policy selected. The tax-transfer function (4′) requires that column 6 equal column 2+0.2 times column 7. The budget restraint (11′) requires that column 3 equal the sum of columns 4, 5, and 6. In each row, 3 policy variables have arbitrarily been chosen as exogenous, and the remaining one (having entries other than 0 or 1 in its column) is dependent.

change in tax rates or government purchases ($\Delta t_0 = \Delta g = 0$ in columns 2 and 3). The high-powered money stock is left to adjust endogenously. In the new equilibrium, NNP is higher by $1.24 billion ($\Delta x = 1.24$ in column 7). At a 20 percent marginal rate this yields induced taxes less transfers of $0.25 billion ($\Delta t = .25$ in column 6). Hence the amount of high-powered money that must be issued in order to complete the financing of the open market purchase, and satisfy the budget restraint, is only $0.75 billion ($\Delta H = .75$ in column 5). Changes in the short-run equilibrium values of the other variables are shown in columns 8–14. Any of the other policies' effects can be read from Table 1 in the same way.

We now compute "multipliers" to express the effects of each monetary and fiscal policy variable upon the short-run equilibrium value of real NNP, for each of the twelve policies. Each multiplier is defined as the quotient of the change in the short-run equilibrium level of real NNP divided by the change in the policy variable in question, thus: $\Delta x/\Delta g$,

TABLE 1—CONTINUED

Effect of the Policy Shown in Columns 2-5 Upon the Ten Dependent Variables (Other than the Dependent Policy Variable)									Linear Combination of 3 Basic Policies (a), (b), (c)
Δt	Δx	Δy	Δz	Δc	$\Delta i=\Delta k$	Δw	Δr	$\Delta r'$	
(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
−.50	2.51	3.01	.35	2.28	.23	1.08	−.00005	−.0037	$a=.50i$
.61	3.06	2.45	.16	1.76	.31	.86	−.00002	−.00005	$b=-.61i+h$
.25	1.24	1.00	1.22	1.16	.09	1.06	−.00017	−.00082	$c=-.25i+j$
−.66	1.69	2.35	−.46	1.51	.17	.37	.00006	.00017	$a-.66c=.66(i-j)$
.48	2.42	1.94	−.47	1.16	.26	.31	.00006	.00037	$b-.52c=-.4-.52j+k$
.33	1.66	1.33	1.63	1.54	.12	141.	−.00023	−.00109	$1.33c=-.33i+1.33j$
1	1.11	.11	−.11	−.01	.13	.02	.00001	.00024	$-.78a+b=-i+k$
−1	2.55	3.55	−.70	2.29	.26	.56	.00010	.00026	$1.51a-c=1.51d=i-j$
−1	5.05	6.05	.70	4.58	.46	2.16	−.00010	−.00074	$2.01a$
0	2.50	2.50	1.40	2.30	.20	1.60	−.00019	−.00100	$.50a+c=i-h$
0	6.16	6.16	.59	4.57	.59	2.18	−.00008	−.00050	$1.23a+b=g+i$
0	3.66	3.66	−.80	2.27	.39	.58	.00011	.00050	$.73a+b-c=g+h=-j+k$

$\Delta x/\Delta t_0$, $\Delta x/\Delta H$, and $\Delta x/30\Delta B$. They are shown in Table 2. (Note that the different multipliers shown for each policy are different ways of expressing the total effect of that policy; to add them together would be double counting, and hence improper.)

For example, in policies a, b, and c the multipliers $\Delta x/\Delta H$ for the high-powered money stock are respectively 5.05, 7.90, and 1.66. The highest and lowest values of each multiplier obtainable among the twelve policies are shown at the foot of each column in Table 2.

We can now summarize the multiplier effects of monetary and fiscal policy variables upon the short-run equilibrium level of real NNP demanded using the foregoing eleven equations in Case 1:

The government purchases multiplier of real NNP may be as great as 6.16 (for policy k which finances the purchases purely by printing high-powered money), and as low as 1.11 (for the balanced-budget multiplier policy g). See column 3.

The high-powered money stock multiplier of real NNP may be as

high as 7.90 (for policy b which uses the additional high-powered money to help finance government purchases, the remaining financing being via induced taxes less transfers), and as low as 1.66 (for policy c which is an open market purchase of government bonds, financed by the additional high-powered money except to the extent that induced taxes less transfers contribute to the financing). See column 5.

The autonomous tax-transfer multiplier of real NNP may be as great as $+1.43$ (for the balanced-budget multiplier policy g), and as low as

TABLE 2

REAL NNP MULTIPLIER, Δ (REAL NNP)$\div\Delta$ (POLICY VARIABLE), IN CASE 1
FOR THE TWELVE SELECTED MONETARY AND FISCAL POLICIES IN TABLE 1

POLICY (1)	REAL NNP MULTIPLIER, $\Delta x/\Delta$ (POLICY VARIABLE), FOR CHANGES IN THESE POLICY VARIABLES:			
	t_0 (2)	g (3)	$30B$ (4)	H (5)
a, i	-2.51	—	—	5.05
b	—	3.06	—	7.90
c, f	—	—	-1.24	1.66
d, h	-1.69	—	2.55	—
e	—	2.42	4.70	—
f, c	—	—	-1.24	1.66
g	1.43	1.11	—	—
h, d	-1.69	—	2.55	—
i, a	-2.51	—	—	5.05
j	—	—	-2.50	2.50
k	—	6.16	—	6.16
l	—	3.66	3.66	—
Max	1.43	6.16	4.70	7.90
Min	-2.51	1.11	-2.50	1.66

SOURCE: Column 7 of Table 1÷column (n) of Table 1 $(n=2, \cdots, 5)$.

-2.51 (for policy a which is a switch from tax finance to printing high-powered money). See column 2.

The private bondholdings multiplier of real NNP may be as great as $+4.70$ (for policy e which issues the bonds in order to help finance government purchases, the remaining financing being via induced taxes less transfers), and as low as -2.50 (for policy j, which is a pure switch between debt finance and printing high-powered money). See column 4.

Next we study the sensitivity of the real NNP multipliers to changes in some of the parameters of the system. We concentrate on interest elasticities of expenditure and asset demand, for it is here that "Keynesian" and "classical" advocates differ most.

For simplicity we consider only three policies, a, b, and c, though others can be examined in the same way.

In Figure 1, the absolute interest elasticity of investment, $|\eta_{ir}|$, varies from the Keynesian extreme of 0 at the left to the strongly classical value of 1 at the right. In computing the NNP multipliers shown in Figure 1, all four interest elasticities of expenditure have been
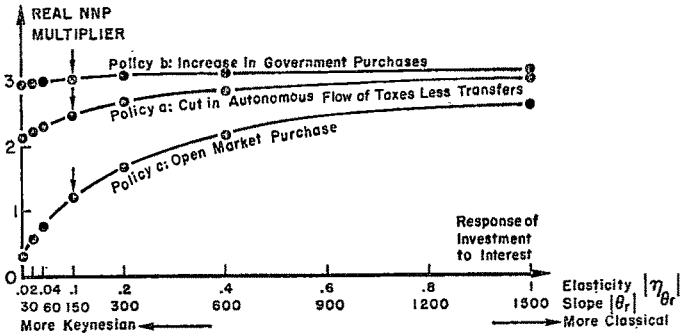
FIGURE 1
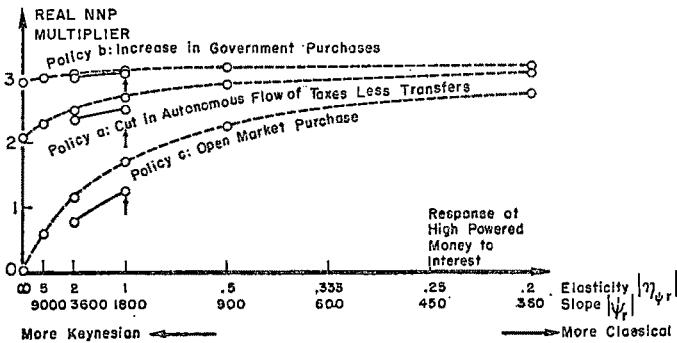
FIGURE 2

varied in proportion to each other, and other parameters have been kept at their Case 1 values.

In Figure 2, the absolute interest elasticity of high-powered money demand, $|\eta_{\psi r}|$, varies from the Keynesian extreme of infinity at the left to the strongly classical value of 0.2 at the right. In computing the NNP multipliers shown by the solid curves in Figure 2, all four interest elasticities of asset demand have been varied in proportion to each other, and other parameters have been kept at their Case 1 values. The dotted curves are the same except that the four interest elasticities of expenditure have been kept at double their Case 1 values.

The two figures show that the real NNP multiplier effects of govern-

ment purchases in policy a and taxes less transfers in policy b are not very sensitive to variations in interest elasticities, and the real NNP multiplier effect of open market operations in policy c is sensitive only toward the Keynesian extremes where expenditure elasticities are low and asset demand elasticities are high. A similar analysis of policies j, k, and l (not shown here) yields the same qualitative result. Hence if we can find a set of parameter values representing the most Keynesian character that the system might realistically have, we can then find a lower bound for the multiplier effect of open market operations. Let us try.

Consider the points in Figure 1 marked by arrows. Here $|\eta_{\theta r}| = 0.1$, as assumed in Case 1, and the absolute investment slope $|\theta_r| = 150$. This would imply that a substantial change of 0.01 in the bond yield r (from $3\frac{1}{3}$ percent to $4\frac{1}{3}$ percent) would cut investment demand by only $1.5 billion, from $50 billion to $48.5 billion. I submit that this is so small as to be a reasonable lower limit, although admittedly we do not have firm econometric evidence on this point.

Consider now the points in Figure 2 marked by arrows. Here $|\eta_{\psi r}| = 1$, as assumed in Case 1, and the absolute high-powered money slope $|\psi_r| = 1800$. This would imply that a change of 0.01 in r (from $3\frac{1}{3}$ to $4\frac{1}{3}$ percent) would cut high-powered money demand by a whopping $18 billion, from $60 billion to $42 billion. I submit that this is so large as to be conservative upper limit, and there is some supporting econometric evidence here; see the references to Christ, Latané, and Meltzer.

If it is accepted that the interest elasticity of investment is no less than 0.1, and that the interest elasticity of high-powered money demand is no more than 1, and that the other assumptions of Case 1 are near the truth, then Figures 1 and 2 show that we can expect a real NNP multiplier of at least $1\frac{1}{4}$ from policy c; that is, from open market purchases financed partly by printing high-powered money and partly by the induced taxes less transfers that arise at existing tax-transfer rates. If open market operations are financed entirely by printing money as in policy j, the multiplier is of course higher, about $2\frac{1}{2}$. These multipliers are too high to justify the neglect of monetary policy.

To summarize: (1) Government policy variables are subject to a budget restraint. Therefore (2) the effect of a change in any single policy variable depends on how other policy variables are varied in order to satisfy the budget restraint.

Therefore (3) the proper object of analysis is not a single policy variable, but a policy (defined as as set of arbitrary values of all but one of the policy variables, the remaining one being determined endogenously by the system). (4) Our rough quantitative results suggest forcefully that both the extreme fiscal advocates and the extreme monetary advocates are wrong: Fiscal variables strongly influence the effect of a

given change in the high-powered money stock, and open market operations strongly influence the effects of given changes in government expenditures and taxation.

REFERENCES AND RELATED WORKS

1. *Phillip D. Cagan, "The Demand for Currency Relative to the Total Money Supply," *J.P.E.*, Aug., 1958, pp. 303–28.
2. *Carl F. Christ, "Interest Rates and 'Portfolio Selection' Among Liquid Assets in the U.S.," in Carl F. Christ and (eleven) others, *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld* (Stanford Univ. Press, 1961), pp. 201–18.
3. Morris A. Copeland, "Tracing Money Flows Through the United States Economy," *A.E.R.*, May 1947, pp. 31–49.
4. *Robert Eisner and Robert Strotz, "Determinants of Business Investment," in Commission on Money and Credit, *Impacts of Monetary Policy* (Prentice-Hall, 1963), pp. 59–337, esp. pp. 151–79.
5. Alain C. Enthoven, "A Neo-classical Model of Money, Debt, and Economic Growth," Mathematical Appendix to *Money in a Theory of Finance*, by John G. Gurley and Edward S. Shaw (Brookings Institution, 1960), pp. 303–59, esp. pp. 315 ff.
6. *Milton Friedman, "The Demand for Money: Some Theoretical and Empirical Results," *J.P.E.*, Aug., 1959, pp. 327–51.
7. Bent Hansen, *The Economic Theory of Fiscal Policy*, translated by P. E. Burke (London, Allen and Unwin, and Cambridge, Harvard Univ. Press, 1958), Chap. 3.
8. *Henry Allen Latané, "Cash Balances and the Interest Rate—A Pragmatic Approach," *Rev. of Econ. and Statis.*, Nov., 1954, pp. 456–60.
9. Assar Lindbeck, *A Study in Monetary Analysis*, Stockholm Economic Studies, New Series III (Stockholm, Almqvist and Wicksell, 1963), pp. 144, 146–47.
10. *Allan H. Meltzer, "The Demand for Money: The Evidence from the Time Series," *J.P.E.*, June, 1963, pp. 219–46.
11. Lloyd Metzler, "Wealth, Saving, and the Rate of Interest," *J.P.E.*, Apr., 1951, pp. 93–116.
12. Richard A. Musgrave, *The Theory of Public Finance: A Study in Public Economy* (McGraw-Hill, 1959), Chap. 22.
13. Don Patinkin, *Money, Interest and Prices* (Row Peterson, 1956), pp. 361–65.
14. *Richard T. Selden, "Monetary Velocity in the United States," in Milton Friedman, ed., *Studies in the Quantity Theory of Money* (Univ. of Chicago Press, 1956), pp. 179–257.

* Note: Starred items contain econometric evidence about elasticities. The other items contain conceptual antecedents to this paper.

DISTRIBUTED LAGS, INTEREST RATE EXPECTATIONS, AND THE IMPACT OF THE MONETARY POLICY: AN ECONOMETRIC ANALYSIS OF A CANADIAN EXPERIENCE*

By Harold T. Shapiro
University of Michigan

The postwar revival of interest in monetary theory and policy remains in full bloom. Increasingly invigorated with new tools of empirical research, all sides are claiming ever fresher and ever more impressive victories. The issue of "sides" itself, however, seems to have become less interesting, if not less fashionable, as careful evaluation and testing of opposing theories seem to have narrowed many of the former differences. Some of the credit for this can be attributed to the increasing influence of quantitative research. Whatever the weaknesses of econometric techniques in the planning and/or evaluating of monetary policy, its methodology and assumptions are clear, and, therefore, open to criticism and adjustment. *Ex post* it is always painfully clear where the mistakes were made and just which ideas are clearly unsupported by the evidence. Even more important, however, they force researchers, not only to develop operational interpretations of theoretical concepts, but also to restrict themselves within this class of conceptually operational concepts to those that can be expressed in terms of accessible data on economic behavior. Perhaps it is the paucity of objective information available that prevents us from declaring clear-cut "winners."

If the instruments of monetary control are to have any influence on real income and employment, they must be able to induce movements in some of the components of either aggregate demand or supply. In addition, if the effects of a particular policy action are to be predictable, there must be stable demand and supply functions for money balances. Initially, therefore, most empirical research centered on the interest elasticity of business investment decisions and on the isolation of a stable demand function for money balances. (As a first approximation, the supply of money was considered under the control of the monetary authorities, and the supply of goods perfectly elastic.) More recently, however, research efforts in this area have moved to the somewhat more complex problem of estimating the time pattern of response of these relations to the instruments of monetary policy, or to those vari-

ables such as interest rates, which these instruments are presumed to affect [1] [8] [12] [13] [14].[1] Obviously, if monetary policy is to be a useful tool of cyclical stabilization, the speed with which it operates on its targets is a crucial factor in such considerations. It is with this problem that the present paper is concerned. In particular, I wish to study the effect of a shift in monetary policy away from a program of pegged bond yields, which occurred along with an important structural reform in the Canadian monetary sector, on the ability of the monetary authorities to use the general instruments of monetary control for short-run stabilization policy.

Recent empirical studies have indicated that various components of aggregate demand do respond to changes in interest rates, but with a lag that generally seems too long for the monetary authorities to make use of this relationship as a vehicle of cyclical stabilization. Tucker [25] has recently pointed out, however, that the existence of long lags in the response of aggregate demand to interest rates is not a sufficient reason for monetary policy to work slowly. Tucker points out that these lags may be counteracted by other lags in the system, particularly by lags in the demand function for money balances. The argument is that, if the demand for money responds only with a lag to a change in the money supply, the interest rate which acts so as to clear this market will temporarily be forced above its new equilibrium level and remain there until the market can adjust to this new disturbance. This exaggerated initial movement in interest rates, Tucker argues, will have a correspondingly exaggerated short-run effect on the components of aggregate demand and, therefore, income. This result implies that despite the long lags in the response of aggregate demand to interest rates, sharp changes in the money supply will still induce correspondingly sharp short-run movements in certain components of aggregate demand.

There are, however, a number of difficulties with Tucker's argument. First, it seems clear to me that the monetary authorities do not plan short-run policy by merely pursuing a particular money supply target irrespective of the associated movements in short-term interest rates. They may, indeed, pursue a money-supply target, but their attempts to achieve it will usually be constrained by certain "acceptable" rates of change in interest rates. This attitude is necessitated by their desire not to "disturb" financial markets. This may simply mean, of course, that this constraint on their actions ought to be removed if we wish to allow monetary policy to operate quickly. More important than this, however, is the simple fact that the short-term interest rate which clears

---

[1] The question of the money supply has now also been extensively investigated [4] [9] [15] [22] [27]. Here attention has centered on the interest elasticity of the demand for excess reserves, in an effort to isolate the leakage between changes in the reserve base and changes in the deposit liabilities of the commercial banks. I will have more to say about this below.

the money market (equates demand and supply of money) is not the interest rate which enters the aggregate demand relations.

In particular, business investment decisions are thought to be a function in part of present and future expected levels of interest rates, a notion which can be allowed for by specifying the long-term interest rate as an argument in this relationship. In general, however, there is no reason to expect that this long-term rate will exhibit the same volatile movements as the short-term rate which may be moving rapidly to accommodate sharp changes in the money supply to the long lags in the demand for money function.[2] In these circumstances (where the amplitude of movements in the long-term rate does not strictly parallel those of the short-term rate) lags in the money demand function will not act as important counteracting forces in the long lags indicated in the aggregate demand relations.

I argue that the key to this controversy lies in the nature of the private sector's expectations regarding future interest rate movements, or to put the matter in a slightly different way, it depends on just how movements in the short-term rate affect the long-term rate. If all movements in short-term rates are immediately transformed into correspondingly sharp movements in long-term rates, then Tucker's argument would hold, and the outlook for monetary policy would be somewhat brighter. If this is not the case, the long lags in the response of aggregate demand to interest rates would still indicate some difficulty in using this relationship as a vehicle of cyclical stabilization policy. Regarding the Canadian situation, I argue that the structural reform in the monetary sector (which is described below) together with a shift away from a monetary policy of fixed bond yields occasioned a marked change in the way expectations of future interest rate levels were formed and thus affected the ability of the monetary authorities—in an unexpected way—to use the instruments of monetary control for short-run stabilization policy.

Section I describes briefly the main characteristics of the Canadian monetary sector and the structural reform on which this study centers attention. Section II develops and estimates the parameters of those relationships necessary to test my hypothesis. In general it is found that, although the structural reform and the sharp accompanying change in monetary policy had no important effects on the demand function for money balances and in particular on the lag structure of this relation, it did, indeed, affect the formation of interest rate expectations in a way that makes it more difficult to use monetary control for cyclical stabilization. It is also argued that the Treasury-Federal Reserve accord may have had a similar effect in the U. S.

---

[2] This follows from the notion that the long-term rate is at least in large part an average of future expected short-term rates (plus a compensating premium) and that there is no reason to expect that all movements in the short-term rate are expected to be permanent.

I. *The Canadian Monetary Sector*

In most respects the Canadian financial system and the general methods of monetary control are very similar to those of the United States. At the center of both private financial systems stand the commercial banks (in Canada referred to as the chartered banks), although the Canadian situation is characterized by extensive branch rather than unit banking. In addition, the Canadian banks play a more dominant role in the overall system of financial intermediation. For our purposes, however, these differences are not significant. The major instruments of general monetary control—the ability to conduct open market operations; the ability to set discount rates and regulate bank borrowing from the central bank; and the ability to set reserve requirements—are the same in both countries. Moreover, both systems have come to place primary reliance on open market operations to carry out the short-run objectives of monetary control.

The ability of the Canadian monetary authorities (the Bank of Canada) to affect the excess demand for money, however, is subject to rather different influences and potential "leakages" between changes in the reserve base and changes in the domestic money supply (broadly defined to include all chartered bank deposit liabilities to the private sector). On the one hand, the Canadian banking system responds to changes in its reserve base by quickly exchanging all excess reserves for earning assets of one type or another. This eliminates leakages into excess reserves. In addition, there is a single reserve requirement on all types of deposit liabilities so that there is no leakage occasioned by the shifting distribution of deposit types.[3] These characteristics tend to simplify the problems of both monetary control and empirical research in this area, as bank portfolio behavior is not such a crucial element in the money supply mechanism. The one exception to this is the potential leakage of excess reserves into foreign assets (largely New York broker loans). Although there is some evidence that chartered bank purchases of foreign assets are sensitive to the Canada-U.S. interest rate differentials, the amounts do not appear to be large or to have any serious effect on the central bank's ability to control the money supply. In other areas, however, the "relative openness" of the Canadian economy considerably complicates the problem of monetary control. Important segments of the Canadian economy are sensitive to developments in foreign markets and to the impact of short- and long-term capital movements. Under a system of fixed exchange rates, monetary policy will be constrained by the obligation to keep fluctuations in the balance of

---

[3] The particular method of calculating reserve requirements also makes it easier for the central bank to set the banking system "scrambling" for reserves toward the end of the accounting period. In any given month the level of excess reserves is under the complete control of the central banks.

payments within the limits set by one's foreign exchange reserves and international borrowing power. A fluctuating exchange rate does not shield the economy from the above influences, but it does permit the monetary authorities more leeway in attempting to prevent these effects from generating fluctuations in aggregate economic activity. In either case, these influences may be responsible for substantial leakages between changes in the reserve base (which is directly under the control of the central bank) and the domestic money supply. For the purposes of this study, I assume that the mechanism governing leakages into international capital flows was stable over the period under investigation (1948-62).

I turn now to the structural reform itself. Beginning in January, 1953, the Bank of Canada, in order to help improve the efficiency of Canadian capital markets in the transfer of funds, introduced a series of reforms designed to encourage the creation of an active short-term money market.[4] These reforms were largely successful and by 1955 a strong and active market for short-term financial instruments had developed. Aside from the emergence of a broad market in short-term government securities, a strong market developed for the short-term obligations of installment finance companies and of prime nonfinancial corporations. Prior to these reforms, the market in private short-term obligations was essentially nonexistent while the market for short-term government securities was largely a contrived affair between the Bank of Canada and the chartered banks. Even after the postwar controls on bond prices had been gradually removed, fluctuations in short-term rates were much more a matter of central bank policy than of changing excess demands for credit.

A bona fide bond market had existed for many years prior to this development and was of critical importance to the central bank in carrying out its operations. The primary objective of introducing this "money-market experiment" was to insure the more efficient transfer and employment of short-term funds, and in this respect, the reforms were highly successful. Once again, however, this study is concerned with the somewhat more restricted question of determining whether these reforms along with a sharp change in the nature of monetary policy affected the lag structure in the demand-for-money function and/or the response of the private sector to changes in the short-term interest rate in a manner that influenced the ability of the central bank to use monetary policy for cyclical stabilization. In particular, I am concerned as to whether these effects were such as to strengthen or weaken the forces counteracting the long lags in the aggregate demand relations.

<hr>

[4] For a more detailed description of the reform see [26].

II. *Lags in the Demand for Money and Interest Rate Expectations*

I will define money as the sum of all commercial bank Canadian dollar deposit liabilities (adjusted to exclude government deposits) and currency in circulation. This corresponds to what is considered the "broad" definition of the money supply. Considering Canadian institutions, however, there is little room for choice in this matter, as savings deposits at commercial banks carry full checking privileges and are used by most households at least in part as transactions accounts. Although there has always been a single reserve requirement against all types of deposit liabilities, these deposits themselves are not a homogeneous quantity, but are made up of demand deposits (45 percent) and personal savings deposits (55 percent).[5] These two types of deposits, however, cannot be assumed to be held for entirely the same reasons and any attempt to isolate the factors which determine the private sector's demand for money balances must deal separately with its three main elements: currency in circulation, demand deposits adjusted, and personal savings deposits.

The variance in the private sector's demand for currency is not yet well understood by economists. Nevertheless, one would certainly expect that the short-run variance in currency holdings is primarily due to its role as a means of payment and not to its role as a form of holding wealth. This being the case, there is little reason to believe that the reform of the monetary sector on which this study centers attention had any effect on desired currency holdings. Parameter estimates of a demand for currency equation revealed that changes in currency holdings were insensitive to changes in interest rates and could largely be explained by movements in income and seasonal factors, but in general responded slowly to changes in the former variable.[6] Moreover, this pattern does not appear to have been altered by innovations in the money market.[7]

In principle, demand deposits should be held only for the purpose of facilitating transactions, and not as a temporary store of wealth [23]. Due to the existence of transactions costs on asset exchanges, however,

---

[5] These figures are for 1962, the last year covered in this study. Since that time, however, *Non Personal Term and Notice Deposits* (an instrument somewhat analogous to the nonnegotiable certificate of deposit in the U.S.) have assumed increased importance. Preliminary investigations seem to indicate that this development has had important effects on the demand for money.

[6] Estimates of the lag structure involved in this relationship indicated that it would take about five quarters for about one-half the effect of a change in the independent variables to make itself felt.

[7] This proposition was tested by generating two independent estimates of the lag structure involved; one based on data generated before the reform and one based on more recent data. The implied lag structures were virtually the same in both cases. More formally, I applied the Chow test [5] to these estimates and could not reject the hypothesis (at the 0.01 level) that the same model generated the data in both periods.

there may be considerable short-run variance in the demand for this asset which is not directly related to current transactions, but related instead to the yield on alternative assets [2] [24]. The 90-day Treasury bill rate is taken as indicative of yields available to business on alternative financial investments. At present, however, there are no data available on the proportion of demand deposits outstanding that is held by business firms. Since personal savings deposits carry checking privileges, however, and are typically used by households as transactions balances, one would expect that the bulk of demand deposits are held by business, or, at very least, that the proportion held by business is much larger than in the United States.

In order to express the essential nonlinear relation between desired money balances and interest rates, I use a logarithmic stock-adjustment model which is outlined below. In addition, velocity (rather than "real" demand balances) is chosen as the dependent variable to eliminate the strong trend in both real money balances and income. The model can be expressed as follows:

$$(1) \qquad \frac{M1_t}{M1_{t-1}} = a_1 \left(\frac{M1_t^*}{M1_{t-1}} \right)^{\gamma_1} U1$$

where:

$M1_t$ = demand deposits at time t (1957 dollars)
$M1_t^*$ = desired deposits at time t (1957 dollars)
γ_1 = speed of adjustment coefficient $0 < \gamma_1 < 1$
$U1$ = stochastic error term

Equation (1) above postulates that a constant proportion (γ_1) of the discrepancy between desired and actual money balances will be eliminated in a single period. The estimation of the speed of adjustment coefficient γ_1 is what is crucial for our purposes, for it contains the information we desire on the length of the lag in the demand-for-money balances. We express the desired level of demand deposit balances in the following fashion:

$$(2) \qquad \ln V1^* = b_1 + b_2 \ln r_{TB}$$

where

$$V1^* = \text{desired velocity} = \frac{Y}{M1^*}$$

$$r_{TB} = \text{90-day Treasury bill rate}$$

This equation implies a number of behavioral assumptions concerning the demand for money. First, it implies a constant interest elasticity of

velocity. Second, it implies that, although economic units will continue to economize on money balances as interest rates rise, they will do so at a decreasing rate. Although velocity also depends on various institutional factors, these are assumed to be constant throughout the period under investigation (1948–62). Taking logarithms of equation (1), substituting (2) into (1) and reformulating the expression in terms of velocity yields the following relationship:

$$(3) \quad \ln V1_t = \beta_1 + \beta_2 \ln r_{TB} + (\beta_3)\Delta \ln Y + (\beta_4) \ln V1_{t-1} + U1$$

where:

$$\beta_3 = \beta_4 = (1 - \gamma_1)$$

Estimating the parameters of this function by ordinary least squares from quarterly observations over the period 1948–62 generates the following estimates:

$$\ln V1_t = 0.199 + 0.013S_1 + 0.038S_2 + 0.067S_3 + 0.062 \ln r_{TB}$$

$$(4.0)^* \quad (.89) \quad (1.8) \quad (2.5) \quad (5.1)$$

(4)

$$+ 0.561\Delta \ln Y + 0.672 \ln V1_{t-1}$$

$$(7.4) \quad (10.6)$$

$$R^2 = 0.977, \text{ S.E.E.} = 0.026, \ln V1 = 0.80, D.W. = 1.98$$

* Figures in parentheses are the t values of the estimated coefficients.

where:

S_1, S_2, S_3 are seasonal dummy variables for the 1st, 2nd, 3rd quarters, respectively

Y = GNP (millions of 1957 dollars—seasonally unadjusted)

r_{TB} = 90-day Treasury bill rate (percent)

The estimated coefficients are all of the expected sign and except for the seasonals are statistically significant at the 0.05 level. In addition, the overall "fit" of the equation is highly satisfactory with an R^2 of 0.977 and a standard error of estimate of 0.026.[8] As indicated above (equation (3)), the estimated coefficients for $\Delta \ln Y$ and $\ln V1_{t-1}$ (β_3 and β_4) yield estimates of the speed of adjustment coefficient γ_1, which in turn determines the lag structure involved in the demand-for-money function. Theoretically the estimated coefficients $\hat{\beta}_3$ and $\hat{\beta}_4$ should yield identical estimates of γ_1 since the model implies that $(1-\beta_3)=(1-\beta_4)=\gamma_1$. This is not quite the case with the above (4), however, as these estimates of β_3 and β_4 yield point estimates of the speed of adjustment (γ_1) of 0.33

[8] The "quality" of these direct least squares estimates will be discussed below.

and 0.44, respectively. Reestimating the model under the constraint $\beta_3 = \beta_4$ yielded a point estimate of 0.39 for γ_1.[9] These estimates imply a substantial lag in the response of velocity to movements in income and interest rates, but a considerably faster response than that typically isolated by studies using U. S. data [9]. This is to be expected, since the proportion of demand deposits held by business firms is much higher in Canada. Thus far, however, I have made no attempt to account for any of the possible statistical biases that may arise from the straight-forward application of ordinary least squares.

The above estimates (4) contain two principal sources of bias. First, there is the simultaneous equation bias which may arise due to the cor-relation of the error term both with r_{TB} and with ΔY. Second, the co-efficient of the lagged dependent variable will be biased especially if there is autocorrelation in the residuals. In the first case (simultaneous equation bias) I am primarily concerned with the effect on the co-efficient of $\Delta \ln Y$ since this coefficient yields estimates of the speed of adjustment coefficient γ_1. It seems clear that, since income will vary positively with the error term, the least squares estimate of β_3 (the income coefficient) will be biased upwards, yielding a conservative (downward biased) estimate of γ_1. Although little is known concerning the small sample properties of simultaneous equation estimators, an attempt was made to remove this bias by estimating the parameters of this relation (3) by two-stage-least squares.[10] The results, however, did not yield point estimates substantially different from the ordinary least squares estimates. The estimated interest rate coefficient was virtually the same in both cases (0.065 vs. 0.067), while the income co-efficient, $\hat{\beta}_3$, now yielded only a marginally higher estimate of γ_1 (0.48 vs. 0.44).

Autocorrelation of the residuals is also a potentially serious source of bias in the ordinary least squares estimate of β_4, the coefficient of the lagged dependent variable. This coefficient is also of critical concern in estimating the lag distribution of the above relationship (3). As ex-pected, the computed Durbin-Watson statistic does not yield any evi-dence of positive autocorrelation, as this statistic is asymptotically biased towards 2 (the value it should have if no serial correlation is present) when lagged dependent variables are specified. Visual exami-nation of the residuals, however, does suggest a substantial amount of

[9] The constrained estimates of this model did not yield as satisfactory an overall "fit" ($R^2 = 0.931$) as the unconstrained case. The estimated interest rate coefficient, however, changed very little—0.071 as compared with 0.067 above.

[10] I obtained the instrumental variables for this procedure by specifying an additional set of equations which together with the above relation (3) determined all the endogenous vari-ables of a highly aggregative economy-wide model. The main instruments were: the domestic monetary base, the U.S. Treasury bill rate, lagged consumption, the balance of trade, govern-ment expenditures, lagged money balances, and lagged GNP. Teigen followed a similar pro-cedure [22].

positive autocorrelation. Our least squares estimates of β_4, therefore, will also be biased upwards, yielding a downward biased or conservative estimate of γ_1.[11] Two procedures were employed in an attempt to remove this source of bias. First, a scheme developed by Hildreth and Lu [11] that yields consistent estimates of the true parameters when the error term $U1_t$ is generated by the following mechanism:

$$U1_t = \lambda U1_{t-1} + V1_t$$

where:

$$|\lambda| < 1, \quad V1\text{'s are serially independent,}$$

was employed. Second, a somewhat modified version of a technique developed by Almon [1] for estimating distributed lags was also tried. This latter technique has the additional advantage of allowing different lag distributions on each of the independent variables. A priori I expected that the response of velocity to interest rates might be slower than the response to changes in income and I allowed for this possibility in employing this procedure. Although the Hildreth-Lu procedure seemed to indicate the presence of some positive autocorrelation (an estimate of $\lambda = 0.4$), neither this technique nor the one proposed by Almon yielded substantially different estimates than the direct least squares procedure of the stock adjustment lag in the money demand relation above (3). Moreover, the evidence indicated that the response patterns of velocity to interest rates and income were very similar, and that they could be adequately represented as an exponential decay.

Now in order to investigate the possible effects of the introduction of the short-term money market on the lag structure of this relationship, two independent estimates of γ_1 were generated: one based on data generated before the reform (1948–52) and one based on the more recent data (1955–62). Although these estimates indicated that the stock adjustment lag was somewhat faster subsequent to the reform, the difference was only marginal, and I could not reject the hypothesis that the same model had generated the data in both periods.[12] A comparison of the estimates from these two subperiods is somewhat complicated by the fact that the problem of autocorrelation in the residuals seems to get worse in the later period (1955–62), thus yielding an even greater upward bias in the estimate of β_4 (the coefficient of the lagged dependent variable). Unfortunately, the Hildreth-Lu procedure was not powerful enough to correct for this. For the moment, therefore, we must content

[11] Griliches [10] has shown that if we have positive autocorrelation in the residuals that follow a first-order Markov system, the bias in the ordinary least squares estimate of the coefficient of the lagged dependent variable will also be positive.

[12] Once again this proposition was tested by applying the Chow test [5]. I could not reject the null hypothesis at the 0.01 level.

ourselves with the conclusion that the lag structure of the relation was not notably altered during the period under investigation.

The remaining component of the money supply, personal savings deposits, is dealt with in very much the same fashion as demand deposits. The same stock adjustment model is used to explain the public's holdings of these deposit liabilities with the exception that the interest rate payable on personal savings deposits is assumed to affect the desired holdings of this financial asset. These considerations yield the following expression:

$$(5) \qquad \ln V_{2t} = e_0 + e_1 \ln r_{TB} - e_2 \ln r_{PS} + e_3 \Delta \ln Y$$
$$+ e_4 \ln V2_{t-1} + U2$$

where:

r_{PS} = interest rate payable on personal savings deposits
$V2 = Y/M2$
$M2$ = personal savings deposits (seasonally unadjusted)
$U2$ = error term
$e_3 = e_4 = (1 - \gamma_2)$
γ_2 = speed of adjustment coefficient (for $M2$)

Estimating the parameters of the above relation (5) from quarterly observations over the period 1948–62 yields the following parameter estimates:

$$(6) \quad \ln V2 = 0.038 - 0.034 S_1 + 0.023 S_2 + 0.056 S_3 + 0.024 \ln r_{TB}$$
$$(1.5) \quad (2.8) \qquad (1.2) \qquad (2.5) \qquad (2.1)$$
$$- 0.054 \ln r_{PS} + 0.597 \Delta \ln Y + 0.871 \ln V2_{t-1}$$
$$(2.1) \qquad\qquad (9.1) \qquad\qquad (15.1)$$
$$R^2 = 0.967, \quad S.E.E. = 0.023, \quad \overline{\ln V2} = 0.204$$

The overall fit is highly satisfactory with all the estimated parameters, save the seasonals, having their expected sign and being statistically significant at the 0.05 level. As might be expected, the estimate of the speed of adjustment coefficient γ_2, generated by the above estimates, is considerably lower than our estimates for γ_1. The point estimate of γ_2, generated by imposing the constraint implied by the above model, $e_3 = e_4$, was 0.22. Once again, however, the constrained estimates did not provide as satisfactory an overall fit ($R^2 = 0.912$). This merely supports the widely held notion that household demands for financial assets will respond somewhat more slowly than that of business firms.

Once again, the same procedures as outlined above in the case of demand deposits to remove the various sources of bias in the ordinary least squares estimates, and to investigate in greater detail the nature

of the lag structure in this relationship and its stability over time were undertaken. As before, however, these procedures did not yield substantially different point estimates of the parameters of this equation (5).

In summary, the above evidence seems to reveal stable demand functions for the various components of the money supply with associated lag structures that appear to have been largely unaffected by the sharp shift in monetary policy in the early 1950's and the associated introduction of the short-term money market but do indicate rather long lags in the response of the demand for money to change in income and interest rates. The lags are, however, considerably longer in the case of personal savings deposits than for demand deposits.

Thus, although the demand for money functions do contain substantial lags which may counteract the interest rate lags in the aggregate demand relations, this component of the process does not seem to have been affected by the monetary reforms of the early 1950's. I will turn now to the other crucial element in the process: the relationship between the long-term and short-term interest rates.

In setting out the relationship between the short-term rate and the long-term rate, I rely heavily on the "preferred habitat theory" recently developed by Modigliani and Sutch [16] [17]. The habitat theory is an adaptation of the conventional expectations theory of the structure of interest rates in a world where: future short-term rates are uncertain, transactors have preferred maturity habitats, and transactors are "risk averters." The theory generates the following relationship:

(7) $$r(m) - r_{TB} = \lambda(m)\Delta r(m)^e + p(m)$$

where:

$r(m)$ = bond yield with m years to maturity

$\lambda(m)$ = proportionality factor

$\Delta r(m)^e$ = "average" expected change in $r(m)$

$p(m)$ = difference between premium commanded by an m period loan and that prevailing in the short market

For purposes of estimation, the expression $\lambda(m)r(m)^e$ can be defined as follows:

(8) $$\lambda(m)r(m)^e_t = c_0(m) + \sum_{i=0}^{m} d_i(m)r_{TB_{t-i}}$$

where the d_i's reflect the way in which recent market experience is extrapolated. The d_i's are a weighting scheme that represents a combination of two elements affecting expectations of future changes in long-term rates, one representing the effect of regressive expectations (expecting the rate to regress towards a "normal" level based on past experience) and the other representing the effect of "extrapolative"

expectations (a rise [fall] in rates lending to an expectation of a further rise [fall]). The weights d_i can roughly be thought of as composed of these two elements in the following fashion, as the difference of two lag structures:

$$(9) \qquad\qquad d_i = a_1 g_i - a_2 x_i$$

where g_i and x_i represent the weights associated with the regressive and extrapolative expectations, respectively. We cannot expect the pattern of the d_i's, therefore, to be of a simple geometric form. Substituting (8) into (7) and rearranging terms yield:

$$(10) \qquad r_{mt} = \delta_0 + \delta_1 r_{TB_t} + \sum_{i=1}^{m} d_i r_{TB_{t-i}} + p(m)_t + U_t$$

where:

U_t = stochastic error term

As a first approximation to the above model, I consider $p(m)$ to be composed of a constant plus a random element which can be merged with U_t.

The high degree of multicolinearity between the independent variables specified rules out straightforward least squares procedures to estimate the parameters of the above relation. Instead, I rely upon a modified version of the Lagrangian interpolation technique recently developed by Almon [1].[13] This procedure estimates the distributed lag coefficients (the d_i's) as points lying on a polynomial which in the present situation we choose to be of the fourth degree. In this case, therefore, the Almon technique requires estimating only five coefficients to generate all the parameter estimates of equation (10).[14]

Employing this technique to estimate the model (10) from the sixty quarterly observations generated over the entire period (1948–62) yielded generally unsatisfactory estimates. The overall fit of the model to the data was fair ($R^2 = 0.81$), but the implied lag distributions were difficult to rationalize. In addition an examination of the residuals revealed a substantial amount of serial correlation in the earlier years, indicating the existence of some further systematic influence at this time. In an effort to improve these estimates and to investigate the effect of the introduction of the short-term money market on the forma-

[13] This technique estimates the coefficients of the distributed lag as points lying on a polynomial which, in our case, is restrained to pass through zero at $l = m$. The technique itself requires little a priori restriction on the lag structure, requiring only that it be approximated by a polynomial.

[14] The coefficients to be estimated are δ_1 and four coefficients associated with the four Almon variables. The actual generation of these Almon variables requires the calculation of a set of Lagrangian interpolation polynomials, which are then used to weight certain past values of r_{TB}. These weighted averages, or Almon variables, are then specified in the direct least squares regression equation.

tion of interest rate expectations, the model was fitted separately to the data generated before and subsequent to this reform. Moreover, this allows some extra flexibility in specifying the length of the lag involved in each case and the degree of the polynomial specified. It proved optimal to specify a second degree polynomial with an eight quarter lag in the earlier period, while a fourth degree polynomial with a sixteen quarter lag proved optimal for the later period. The following estimates were obtained:

(11)   1948–54 $r_L = 2.557 + 1.241 r_{TB} + \sum_{i=1}^{8} d_i r_{TB_{t-i}}$

$\qquad\qquad\quad$ (2.7)   (5.5)

$\qquad\qquad$ $R^2 = .831, \quad S.E.E. = .15, \quad \sum_i d_i\text{'s} = -0.31$

(12)   1955–62 $r_L = 1.851 + 0.228 r_{TB} + \sum_{i=1}^{16} d_i r_{TB_{t-i}}$

$\qquad\qquad\quad$ (2.1)   (7.4)

$\qquad\qquad$ $R^2 = 0.975, \quad S.E.E. = 0.11, \quad \sum_i d_i\text{'s} = 0.693$

where:
r_L = interest rate on government bonds with over 10 years to maturity
The summation terms

$$\left(\sum_i d_i r_{TB_{t-i}} \right)$$

represent the distributed lag in each case and contain the information we desire concerning the formation of interest rate expectations (see (9) above).

It is apparent from the above statistics that the model fits the data remarkably well in both periods. The estimates of the lag structure (the d_i's) are plotted in Figures 1 and 2.[15] The implied lag structures from each of these estimates, however, are strikingly different. The structure plotted in Figure 2, together with the estimated coefficient on the current short-term rate, can be interpreted [16] as providing support for the hypothesis that expectations for future interest rates involve significant extrapolative as well as regressive elements. This pattern is very similar to that isolated from the U.S. data by Modigliani and Sutch for the post-1952 period (post-accord period). The structure implied for the earlier period (Figure 1), however, implies a much stronger prevalence of extrapolative expectations as well as a shorter overall lag.

[15] For the 1948–54 period, seven of the eight estimated coefficients describing the lag structure are statistically significant at the 0.05 level. For the later period, twelve of the sixteen coefficients are statistically significant at the 0.05 level.
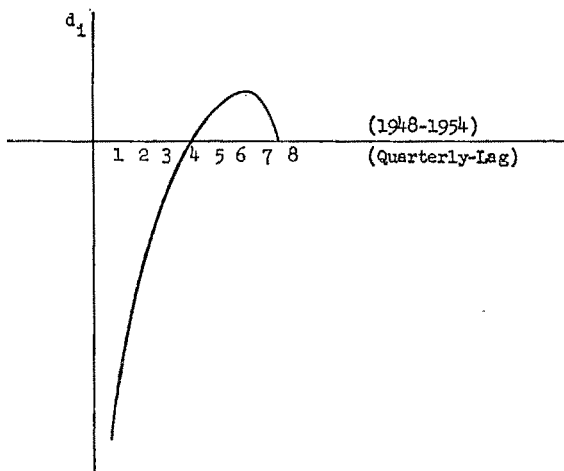
FIGURE 1
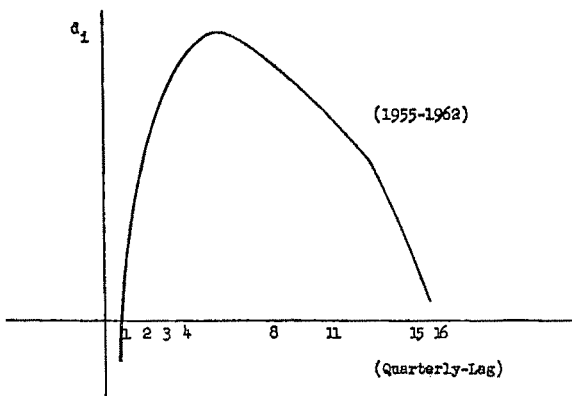
FIGURE 2

The estimated coefficient for the current short-term rate is now some-what larger than one (1.24), strongly indicating that movements in short-term rates are quickly translated into equivalent movements in long-term rates. Moreover, the estimated relative weights attached to the current versus past levels of short-term rates (i.e., δ_1 versus $\sum_i d_i$) is now much larger yielding further evidence of the dominance of extrapolative expectations during this earlier period.

This, perhaps, is the rational response in a market where movements in short-term interest rates represent not the varying excess demand and supply for credit, but rather specific policy actions of the central bank.

In the earlier years of this period (1948–51) monetary authorities were still committed to a modified program supporting bond prices. In these circumstances it is not surprising that investors expected movements in interest rates to be "permanent" or to be followed by further moves in the same direction. Response to movements in short-term rates was similar to the response that might be expected, for example, to changes in the Federal Reserve's discount rate. In this earlier period, changes in interest rates were taken as an indicator of important changes in central bank policy.

If the above intepretation is correct, one should be able to isolate a similar shift in interest rate expectation patterns in the U.S. if we compare the pre-accord and the post-accord periods. As noted above, Modigliani and Sutch [16] [17], using U.S. data, have isolated a pattern, similar to one revealed from Canadian data in the period 1955–62, for the period since 1952. The pre-accord period, however, was largely one of controlled bond prices and it is reasonable to expect that investors would exhibit a strongly extrapolative expectation pattern as was the case in Canada prior to 1954.[16] The reasoning would be precisely the same as in the Canadian case and merely reflects the notion that investors would interpret changes in the short-term rates as reflecting important policy actions of the central bank, and, therefore, would expect such movements to be permanent or to be continued in the same direction. Fitting the model to U.S. data for the period 1946–51 does reveal the expected shift in the expectations pattern. Similar to the Canadian case, it proved optimal to specify a second degree polynomial with a ten quarter lag for this period. The following estimates were obtained:

$$r_{LUS} = \underset{(5.5)}{2.013} + \underset{(5.9)}{0.581r_{TBUS}} + \sum_{i=1}^{10} d_i r_{TB_{t-1}}$$

$$R^2 = .847, \quad S.E.E. = 0.07, \quad \sum_i d_i\text{'s} = -0.21$$

where:

r_{LUS} = interest rate on long-term government bonds (U.S.)
r_{TBUS} = Treasury bill rate—90 day, U.S.

The estimated d_i's are plotted in Figure 3. This pattern is remarkably similar to that isolated from the Canadian data for the pre-reform period. Again I interpret the relatively more important weight attached to the

---

[16] In this connection it is not surprising that the Federal Reserve System only began to revive the discount mechanism subsequent to the accord. Prior to this, movements in bond prices conveyed the same "psychological" effects. For a further discussion of this see [20].

current versus lagged levels of short-term rates as reflecting strongly extrapolative expectations.[17]

Returning to our main theme, it is clear that the long lags in the demand for money relations act as a counteracting force to the equally long lags in the aggregate demand relations only if changes in short-term rates are quickly translated into equivalent movements in long-term rates. This seems to have been the case both in the U.S. and in Canada during the late 1940's and early 1950's. We have explained this
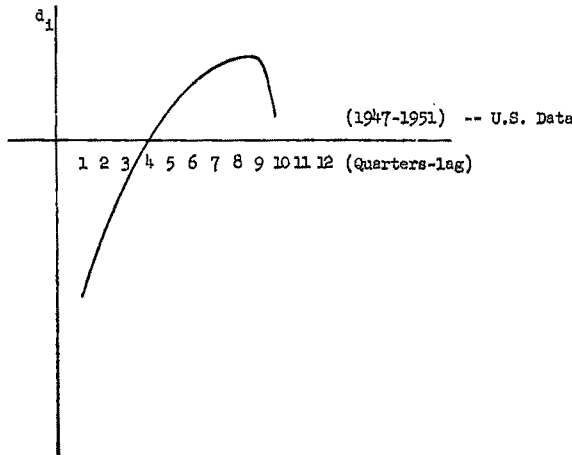
FIGURE 3

type of interest rate behavior in terms of the "expectations" theory of interest rates. If this is not the case, however, these lagged responses will not offset each other, and there is little the monetary authorities can do to overcome this delayed response of aggregate demand to the instruments of monetary policy. In the Canadian situation, the introduction of the short-term money market seems to have had the effect of shifting the interest rate expectations mechanism from a strongly extrapolative scheme to one which combined both extrapolative and regressive elements. The same seems to be true for the U.S. when we compare the pre-accord and post-accord periods. The effect of this shift is to buffer the long-term rate from the more volatile movements of the Treasury bill rate, and thus weaken the forces that may counteract the lagged response of aggregate demand to the instruments of monetary controls.

[17] Although Modigliani and Sutch [16] report that the preferred habitat model as formulated above (10) can account for 95 percent of the variance in the long-term rate, the above considerations suggest that, if some attempt were made to incorporate the effects on interest rate expectations of change in the discount rate, a significant proportion of the remaining 5 percent may also be explained.

REFERENCES

1. S. Almon, "The Distributed Lag Between Capital Appropriations and Expenditures," *Econometrica*, Jan., 1965.
2. W. J. Baumol, "Transactions Demand for Cash: An Inventory Theoretic Approach," *Q.J.E.*, Nov., 1952.
3. *Bookings Quarterly Econometric Model of the United States Economy*, J. Duesenberry, G. Fromm, L. Klein, E. Kuh, eds. (Rand McNally and North-Holland Pub. Co., 1965).
4. K. Brunner and A. Meltzer, "Some Further Investigations of Demand and Supply Functions for Money," *J. of Fin.*, May, 1964.
5. G. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica*, 1960.
6. F. de Leeuw, "A Model of Financial Behavior," Chap. 13 in [3].
7. W. G. Dewald, "Free Reserves, Total Reserves, and Monetary Control," *J.P.E.*, Apr., 1963.
8. M. Friedman, "The Demand for Money: Some Theoretical and Empirical Results," *J.P.E.*, Aug., 1959.
9. S. M. Goldfeld, *Commercial Bank Behavior and Economic Activity* (North-Holland Pub. Co., 1966).
10. Z. Griliches, "A Note on Serial Correlation Bias in Estimates of Distributed Lags," *Econometrica*, Jan., 1961.
11. C. Hildreth and J. Lu, "Demand Relations with Autocorrelated Disturbances," *Michigan State Univ. Agric. Exp. Sta. Dept. of Agric. Econ.* (1960).
12. H. G. Johnson, *The Canadian Quandary* (McGraw-Hill, Toronto, 1963).
13. H. G. Johnson and J. W. Winder, *Lags in the Effects of Monetary Policy in Canada*, Staff study for the Royal Commission on Banking and Finance (Nov., 1962).
14. J. Kareken and R. Solow, "Lags in Monetary Policy," in Commission on Money and Credit, *Stabilization Policies* (Prentice-Hall, 1963).
15. A. J. Meigs, *Free Reserves and the Money Supply* (Univ. of Chicago Press, 1962).
16. F. Modigliani and R. Sutch, "Innovations in Interest Rate Policy," *A.E.R.*, May, 1966.
17. ———, "Debt Management and the Term Structure of Interest Rates: An Empirical Analysis of Recent Experience," Conference of University Professors, The American Bankers Assn. (mimeo., Sept. 19, 1966).
18. M. Nerlove and K. F. Wallis, "Use of the Durbin-Watson Statistic in Inappropriate Situations," *Econometrica*, Jan., 1966.
19. H. T. Shapiro, "The Canadian Monetary Sector: An Econometric Analysis" (unpublished Ph.D. dissertation, Princeton Univ., 1964).
20. W. L. Smith, "The Discount Rate as a Credit Control Weapon," *J.P.E.*, Apr., 1958.
21. R. L. Teigen, "A Structural Approach to the Impact of Monetary Policy," *J. of Fin.*, May, 1964.
22. ———, "Demand and Supply Functions for Money in the United States: Some Structural Estimates," *Econometrica*, Oct., 1964.
23. J. Tobin, "Liquidity Preference as Behavior Towards Risk," *Rev. of Econ. Studies*, Feb., 1958.
24. ———, "The Interest Elasticity of the Transactions Demand for Cash," *Rev. of Econ. and Statis.*, Aug., 1956.
25. D. P. Tucker, "Dynamic Income Adjustment to Money Supply Changes," *A.E.R.*, June, 1966.
26. J. S. G. Wilson, "The Canadian Money Market Experiment," *Banca Nazionale Del Lavoro*, Mar., 1958.
27. J. Wood, "A Model of Federal Reserve Behavior," *Staff Economic Studies* (Board of Governors of the Federal Reserve System, 1966).

EXPECTATIONS AND ADJUSTMENTS IN THE MONETARY SECTOR*

By EDGAR L. FEIGE

University of Wisconsin

The growing interest in the application of distributed lag models has important implications for the area of monetary economics. Although distributed lag models have been most extensively used in connection with studies of consumption and investment expenditure behavior, they also appear to have very attractive features for specifications of financial behavior. The theoretical underpinnings of distributed lag models have in general been quite weak, and as Griliches [8] has pointed out in his recent survey on distributed lags, these models are often open to the charge of "theoretical ad-hockery." The two important exceptions to this generalization are distributed lag models which are rationalized by reference to an adaptive expectation mechanism or a partial adjustment mechanism. Since monetary theory places particular emphasis on the role of expectations in portfolio behavior and on speeds of adjustment to desired positions, it seems natural that empirical studies would come to focus greater attention on the manner in which expectations are formulated and the manner in which both expectations and adjustments to desired positions affect the monetary sector.

The purpose of this paper is, first, to provide a rough sketch of the theoretical underpinnings of the adaptive expectation and partial adjustment mechanisms as applied to the monetary sector. We then analyze the problem of empirically discriminating between adaptive expectation and partial adjustment processes and finally present some preliminary estimates of the structural parameters of a simplified monetary sector which incorporates both adaptive expectation and partial adjustment processes.

Simple Adaptive Expectations and Partial Adjustments

A number of econometric studies of the demand for money have specified demand functions which include the lagged value of cash balances as an argument of the demand for money.[1] In order to properly interpret such studies, it is necessary to know the explicit structure of

[1] See Bronfenbrenner and Mayer [1], Chow [3], DeLeeuw [4], and Teigen [15].

the monetary sector so that one can interpret the coefficients of the reduced equations in terms of the stuuctural parameters of the model.

The difficulty in interpreting the coefficients from reduced equations results from the fact that a number of alternative structural specifications will give rise to similar reduced equations.[2] In order to illustrate this point we begin by considering a simple model of the monetary sector which incorporates both adaptive expectation and partial adjustment mechanisms.

The model specifies that the "desired" long-run stock of real cash balances depends upon "expected" real income such that

$$(1) \qquad\qquad m_t^* = \beta y_t^e + u_t$$

where

> m_t^* is the "desired" long-run stock of real cash balances
> y_t^e is "expected" real income
> u_t is a disturbance term

and lower case letters are used to denote logarithms throughout the paper.

Since both the desired long-term stock of real cash balances and expected real income are theoretical magnitudes which are not directly observable, it is necessary to relate these theoretical magnitudes to observed values. This can be accomplished by first specifying a partial adjustment process which relates current effective demand for real cash balances to the long-run desired stock such that

$$(2) \qquad\qquad m_t^d = m_{t-1} + \gamma(m_t^* - m_{t-1}),$$

where m_t^d is the current effective demand for real cash balances and γ is simply the adjustment elasticity.

Expected real income can also be related to observable values by specifying an expectation generating equation such as,

$$(3) \qquad\qquad y_t^e = y_{t-1}^e + \lambda_y(y_t - y_{t-1}^e)$$

which implies that income expectations are revised in proportion to the proportionate error associated with previous levels of expectations.[3] Since equation (3) is linear in logarithms, λ_y is simply the elasticity of income expectations. Equation (3) can be solved for expected income as a function of all past values of realized income

---

[2] Griliches [8], Waud [17], and Zellner [18] all contain excellent discussions of the difficulties involved in interpreting distributed lag formulations.
[3] See Nerlove [14].

$$(3') \qquad \overset{e}{y_t} = \lambda_y[y_t + (1 - \lambda_y)y_{t-1}$$
$$+ (1 - \lambda_y)^2 y_{t-2} + \cdots + (1 - \lambda_y)^n y_{t-n} + \cdots]$$

and in this form, the expectation generating equation can be viewed as a distributed lag with geometrically declining weights. The specification of the monetary sector is completed by assuming that the supply of real cash balances depends upon the predetermined value of nominal cash balances and the price level, such that,

$$(4) \qquad \qquad \overset{s}{m_t} = z_t - p_t$$

where

m^s_t is the real supply of cash balances
z_t is the predetermined nominal money supply
p_t is the price level

and the equilibrium condition for the monetary sector is

$$(5) \qquad \qquad \overset{d}{m_t} = \overset{s}{m_t} = m_t$$

where m_t is the observed stock of real cash balances. Specifications similar to (4) and (5) are indeed implicit in those models which simply relate the observed stock of real cash balances to variables which are assumed to affect the demand for money. Combining equations (1)–(5) and applying the Koyck transformation results in the following reduced equation for the monetary sector[4]

$$(6) \qquad m_t = [(1 - \lambda_y) + (1 - \gamma)]m_{t-1} - [(1 - \lambda_y)(1 - \gamma)]m_{t-2}$$
$$+ \gamma\lambda_y\beta y_t + \gamma[u_t - (1 - \lambda_y)u_{t-1}].$$

When only the adaptive expectation process is operative, that is when $\gamma = 1$ and $0 < \lambda_y < 1$, the reduced equation (6) becomes

$$(7) \qquad m_t = (1 - \lambda_y)m_{t-1} + \lambda_y\beta y_t + u_t - (1 - \lambda_y)u_{t-1}.$$

Alternatively, when only the partial adjustment process is operative, namely when $\lambda_y = 1$ and $0 < \gamma < 1$, the reduced equation becomes

$$(8) \qquad \qquad m_t = (1 - \gamma)m_{t-1} + \gamma\beta y_t + \gamma u_t.$$

Since equations (7) and (8) contain the same observed variables, one cannot discriminate between the simple adaptive expectation process and the partial adjustment process, except insofar as the disturbance terms differ. When both processes are operative, the reduced equation (6) is relevant; however, one cannot, on the basis of knowledge

[4] See Koyck [9].

of the reduced equation coefficients, identify λ_v, γ and β separately. It is clear, however, that when both processes are operative, the reduced equation for the monetary sector must include two lagged values of cash balances. Thus the dynamic properties of the monetary sector depend critically on the initial specification of the structure.

The adaptive expectation and partial adjustment processes can be derived from a capital theoretic interpretation of the monetary asset. In this perspective, money is regarded as the financial analogue of a consumer durable good which is held primarily for the flow of services yielded by the durable. Friedman's extension of the permanent income hypothesis to the demand for money suggests that wealth, or, alternatively, the expected yield on wealth, is the most important determining variable affecting the demand for cash balances.[5] Expected income is assumed to be an appropriate proxy for the expected yield on wealth, and the expectation generating equation utilized above is analogous to Friedman's permanent income construct.[6] The implication which Friedman derives from his analysis is that money does not perform the role of shock absorber or buffer stock in the portfolio of individuals but rather, the short-run buffer stock function is relegated to "other balance sheet items such as personal debt, consumer credit and perhaps securities." This view is generally regarded in contradistinction to the traditional notion that the transactions motive for holding cash balances is dominant, and thus, that the demand for cash balances is determined by measured income or, more generally by one's forecast of measured income. Perhaps a more direct interpretation of the adaptive expectation model is to regard expected income as a forecast of measured income. Muth [12] has demonstrated that if the process generating measured income is such that the change in measured income is a first order moving average of random deviates, then the expectation generating function described by equation (3) provides an optimal forecast of measured income. In the case where the elasticity of income expectations is unity, the demand for money is simply a function of current measured income. Thus, the adaptive expectation mechanism can be interpreted as representing either an optimal forecast of measured income, or, alternatively, a function which generates a proxy for the expected long-term yield on wealth.

Given the adaptive expectation framework, one can also provide a rationalization for the particular partial adjustment process described by equation (2).[7] Assume that an individual, given his expected income,

[5] See Friedman [7].
[6] See Friedman [6].
[7] The following analysis is analogous to the development in Griliches [8] which considers a rationalization of a partial adjustment process in a capital accumulation framework. The original development is to be found in Eisner and Strotz [5] in connection with the theory of investment expenditures.

chooses a long-run desired level of cash balances (m_t^*). We now wish to consider the costs associated with a particular cash balance position (m_t). The costs can be broken down into two components: (1) the cost associated with being out of long-run equilibrium and (2) the direct costs of portfolio change The cost of being out of equilibrium is assumed to depend upon the gap between the individual's current cash position and his long-run desired position. If his current cash position exceeds his long-run desired position, the individual suffers the cost of foregone income. Alternatively, if his current position falls short of his long-run desired position, he suffers the costs of increased risk and inconvenience. We can represent this cost by

(9) $$c_1 = \alpha(m_t - m_t^*)^2.$$

The second cost incurred represents the brokerage charges and other transaction costs associated with changes in the portfolio, and these costs are assumed to depend upon the change in the current cash position. Thus

(10) $$c_2 = \delta(m_t - m_{t-1})^2.$$

The cost functions are assumed to be quadratic, and the total cost function can be written as

(11) $$c = c_1 + c_2 = \alpha(m_t - m_t^*)^2 + \delta(m_t - m_{t-1})^2.$$

The problem, then, is to choose that cash position (m_t), which, given the long-run desired position (m_t^*) and the previous cash position (m_{t-1}), minimizes total cost. Differentiating (11) with respect to current cash position and setting the derivative equal to zero enables one to solve for the current cash position, such that

(12) $$m_t = \frac{\alpha}{\alpha + \delta} m_t^* - \frac{\delta}{\alpha + \delta} m_{t-1}.$$

Defining $\gamma = \alpha/\alpha + \delta$ and rearranging terms, results in the following expression for the current cash position as a function of the long-run desired position and the previous cash position:

(13) $$m_t = m_{t-1} + \gamma(m_t^* - m_{t-1}).$$

Equation (13) is simply the partial adjustment mechanism specified in the original model. The elasticity of adjustment will depend upon the marginal costs of being out of equilibrium relative to the marginal costs of portfolio change. In this connection it is interesting to note that if the major impact of innovation in financial intermediation is to reduce

the costs of portfolio change by reducing the spread between borrowing and lending rates, we can expect more rapid adjustment of actual cash balances to desired positions. Alternatively, if the major impact of financial innovation is to provide close substitutes for money, thus reducing the costs associated with a disequilibrium position, the effect may be to reduce the speed of cash balance adjustments.

Multiple Expectations and Partial Adjustments

The simple model presented in the preceding section has two important drawbacks. The first of these is the omission of expected interest rates in the demand function for real cash balances and the second the empirical impossibility of discriminating between the simple adaptive expectation mechanism and the partial adjustment mechanism. Both inadequacies can be remedied by introducing the expected rate of interest into the demand function, in order to represent the substitution possibilities available to holders of cash balances. We thus replace equation (1) by

(1')
$$m_t^* = a + \beta_1 y_t^e + \beta_2 r_t^e + u_t$$

where, r_t^e is the expected rate of interest which is generated by the adaptive expectation mechanism

(14)
$$r_t^e = r_{t-1}^e + \lambda_r(r_t - r_{t-1}^e),$$

and λ_r is the elasticity of interest rate expectations. This expectation generating function is again subject to two interpretations. The expected interest rate can either be viewed as an optimal forecast of the short-term interest rate or alternatively can be interpreted as an average long-term rate expected in the future.

The foregoing formulation of the monetary sector is of considerable interest insofar as it takes explicit account of both interest rate and income expectation effects in addition to a dynamic specification of the process whereby actual cash balance positions are adjusted to desired levels. A special case of this model has been widely used in estimating the demand function for money. The special case is derived by constraining the elasticities of expectation and adjustment to be equal to unity and thus to define the demand function solely in terms of current interest rates and current income. Other studies which have directly employed Friedman's constructed permanent income series (derived from his consumption function study) can be regarded as implicitly constraining the elasticity of income expectations to equal 0.40 which is Friedman's estimate, and furthermore, constraining the interest expectations elasticity to equal unity.[8] The advantage of the present

[8] See for example Chow [3], Laidler [10], and Meltzer [11].

model is that it allows one to simultaneously derive estimates of the behavioral elasticities, the expectations elasticities and the adjustment elasticity without the imposition of prior constraints on the values of these parameters.

When equations (1')–(5) are combined with the expectation generating equation for interest rates, one can, after repeated application of the Koyck transformation, derive the following reduced equation for the monetary sector

$$(15) \qquad m_t = c_0 + c_1 m_{t-1} + c_2 m_{t-2} + c_3 m_{t-3} + c_4 y_t$$
$$+ c_5 y_{t-1} + c_6 r_t + c_7 r_{t-1} + v_t.$$

The coefficients of the reduced equation (15) are simply functions of the structural parameters of the model and the disturbance term of the reduced equation can be expressed in terms of the original disturbances as

$$(16) \qquad v_t = u_t - \theta_1 u_{t-1} + \theta_2 u_{t-2}$$

where the θ's are also functions of the structural parameters. In terms of equation (15) which contains eight variables, the six structural parameters are overidentified. Therefore, estimation of (15) by constrained nonlinear least squares is required in order to obtain unique estimates of the structural parameters. It is possible, then, to estimate all of the structural parameters simultaneously, and thus to avoid the difficulty of lack of identification encountered in the simple adaptive expectation and partial adjustment model.

The nonlinear least squares estimates of the structural parameters would, however, be subject to two principal sources of bias; namely, simultaneous equation bias and bias due to the possible presence of autocorrelation in the disturbance term of the reduced equation. Since the monetary sector which we have specified represents only a segment of a simultaneous equation model in which real income, the rate of interest and the price level are all endogenously determined, one can avoid the problem of simultaneous equation bias by utilizing a two-stage estimation procedure.[9]

In order to deal with the problem of autocorrelated disturbances we postulate that the reduced form disturbance assumes the following form

$$(17) \qquad v_t = \rho v_{t-1} + \epsilon_t$$

where ρ is the autoregressive coefficient and ϵ_t is a non-autocorrelated error term with zero mean. In terms of the original disturbances, we must assume that they follow a rather general third order auto-

<hr>

[9] The instrumental variables employed are high-powered money, the discount rate, government expenditures and exports.

regressive process; namely,

(18) $\qquad u_t = (\rho + \theta_1)u_{t-1} - (\theta_2 + \rho\theta_1)u_{t-2} + \rho\theta_2 u_{t-3} + \epsilon_t.$

When specification (17) is combined with the other equations of the model, the resulting reduced equation is

(19) $\qquad m_t = d_0 + d_1 m_{t-1} + d_2 m_{t-2} + d_3 m_{t-3} + d_4 m_{t-4} + d_5 y_t$
$$+ d_6 y_{t-1} + d_7 y_{t-2} + d_8 r_t + d_9 r_{t-1} + d_{10} r_{t-2} + \epsilon_t.$$

The coefficients of (19) can be expressed in terms of the structural parameters of the model as follows:

$$d_0 = \gamma \lambda_y \lambda_r (1 - \rho)a$$
$$d_1 = (3 - \gamma - \lambda_y - \lambda_r + \rho)$$
$$d_2 = -[(3 - \gamma - \lambda_y - \lambda_r)\rho + (1 - \gamma)(1 - \lambda_y)$$
$$+ (1 - \gamma)(1 - \lambda_r) + (1 - \lambda_y)(1 - \lambda_r)]$$
$$d_3 = \rho[(1 - \lambda_y)(1 - \lambda_r) + (1 - \gamma)(1 - \lambda_r) + (1 - \gamma)(1 - \lambda_y)]$$
$$+ (1 - \gamma)(1 - \lambda_y)(1 - \lambda_r)$$
$$d_4 = -(1 - \gamma)(1 - \lambda_y)(1 - \lambda_r)\rho$$
$$d_5 = \gamma \beta_1 \lambda_y$$
$$d_6 = -\gamma \beta_1 \lambda_y [\rho + (1 - \lambda_r)]$$
$$d_7 = \gamma \beta_1 \lambda_y (1 - \lambda_r)\rho$$
$$d_8 = \gamma \beta_2 \lambda_r$$
$$d_9 = -\gamma \beta_2 \lambda_r [\rho + (1 - \lambda_y)]$$
$$d_{10} = \gamma (1 - \lambda_y) \beta_2 \lambda_r \rho.$$

Since equation (19) contains ten variables, the seven structural parameters are overidentified. Estimation of equation (19) by a constrained nonlinear two-stage estimation procedure enables us to obtain unique estimates of all of the structural parameters simultaneously.[10] The results from the constrained two-stage estimation procedure are reported in Table 1 for money defined as currency plus demand deposits (M_1), and for money defined as currency plus demand deposits plus time deposits (M_2).[11]

The estimated income elasticity, β_1, for the narrow definition of money is about 1.3 and does significantly differ from unity, whereas

[10] See Zellner [18].
[11] The data utilized in this study are annual observations covering the period 1915–63. The income series is Kuznets' estimate of net national product and the price series is the implicit NNP price deflator. The interest rate is the commercial paper rate. The data were generously supplied by Anna Schwartz and Milton Friedman.

TABLE 1

CONSTRAINED TWO-STAGE LEAST SQUARES ESTIMATES OF THE PARAMETERS OF THE
MULTIPLE ADAPTIVE EXPECTATION—PARTIAL ADJUSTMENT MODEL
1915–63

Quantity Estimated	M_1		M_2	
	Coefficient	Standard Error	Cofficient	Standard Error
β_1	1.282	(0.081)	1.073	(0.084)
β_2	−0.195	(0.036)	−0.136	(0.034)
λ_y	0.373	(0.087)	0.296	(0.065)
λ_r	0.745	(0.276)	0.854	(0.291)
γ	1.228	(0.375)	1.091	(0.399)
ρ	0.165	(0.535)	0.239	(0.489)
a	−0.367	(0.304)	1.000	(0.111)

the income elasticity for Friedman's broader definition of money does not differ significantly from unity.[12] As expected, the interest elasticities, β_2, are negative and significantly different from zero, thus rejecting Friedman's contention that once expected income is included in the demand function that interest rates no longer contribute to explaining variations in cash balances. This result can be given two interpretations. First, if the yield on wealth is the appropriate scale variable in the demand function for money, then expected income may be regarded as a poor proxy for the expected yield on wealth, and the interest rate effect can be interpreted as reflecting additional information on the expected wealth position. Alternatively, and I believe more convincingly, one could accept the interpretation that the transaction motive is appropriately captured by expected income, viewed as an optimal predictor of measured income, and that the independent effect of the interest rate simply reflects the relevant substitution possibilities confronting the holder of cash balances. This latter view appears to be supported by the finding that the interest elasticity for M_1 exceeds the interest elasticity for M_2. This result is not surprising since if the interest rate utilized is correlated with the yield on time deposits, the computed interest elasticity for the broader definition of money will reflect both own and cross elasticities and thus will be biased toward zero.

Of particular interest are the estimates of the expectation and adjustment elasticities. The estimates of the income expectation elasticities, · λ_y, are 0.37 and 0.30 for M_1 and M_2, respectively. These results represent independent evidence to support Friedman's contention that the constructed permanent income weights derived from his study of the consumption function might be directly applicable to the demand for money since his computed estimate of the income expectations elasticity is 0.40.[13]

The interest rate expectation elasticities do not significantly differ

[12] The 0.05 level was adopted for all tests of significance.
[13] See Friedman [7].

from unity, suggesting that interest rate expectations are static, or, alternatively, that current interest rates appropriately reflect the substitution possibilities available to holders of cash balances.

Of considerable interest is the estimated elasticity of adjustment which does not significantly differ from unity for either definition of money.[14] The unitary elasticity suggests that cash balance portfolio adjustments to desired positions are completed within a single year. In terms of the cost minimization model of the partial adjustment process previously presented, the result suggests that the costs of being out of equilibrium with respect to cash balances far outweigh the costs of portfolio adjustment.

The estimates of the structural parameters suggest that some concept of expected income rather than current income ought to be used in specifying the demand for cash balances; however, they cast doubt on the usefulness of extending the expectation framework to interest rates, even though expected interest rates have received considerable theoretical attention.[15] The analysis also suggests an alternative to utilizing the "permanent" weights derived by Friedman, insofar as the "expectation" weights can be derived simultaneously with the other parameters of the model.

By separating expectation elasticities and adjustment elasticities it is now possible to identify the process which generates lagged values of cash balances in the reduced equation for the monetary sector. Previous studies of the demand for money have found significant coefficients for lagged money balances and have concluded on this basis that there exist substantial lags in the adjustment of actual cash balances to desired positions.[16] The foregoing results suggest an alternative interpretation; namely, that the coefficient of lagged cash balances reflects the effects of expected income rather than a partial adjustment to desired positions. This can be most easily seen by substituting the estimated values of the structural parameters into the expressions for the coefficients of the reduced form equation (19). Since λ_r and γ are approximately equal to unity, and the autoregressive parameter ρ does not significantly differ from zero, equation (19) reduces to

$$(20) \qquad m_t = (1 - \lambda_y)m_{t-1} + \beta_1\lambda_y y_t + \beta_2 r_t - \beta_2(1 - \lambda_y)r_{t-1}.$$

The coefficient of lagged cash balances now simply reflects the effects of the adaptive expectation process relating to income expectations. From the macro viewpoint, we must attach some significance to the

[14] Mundlak [13] has demonstrated that aggregation over time can give rise to biased estimates of the adjustment elasticity; however, his analysis suggests that estimates based on annual data are likely to be underestimates of the true adjustment elasticity.

[15] The author has also estimated a model which includes expected prices and population in the demand function for money; however, neither expected prices nor population appears to have a significant effect on the demand for cash balances.

[16] See Chow [3], Teigen [15], and DeLeeuw [4].

finding that the coefficient of lagged balances is nonzero, since Tucker [16] has demonstrated that the existence of nonzero coefficient for lagged cash balances can offset lagged responses in the expenditure sector, which would otherwise impair the speed with which monetary policy can affect the level of income and employment.

Summary

The application of distributed lag models to the monetary sector appears to be a highly useful tool for investigating the effects of expectations on financial behavior and for analyzing the dynamic adjustment properties of the financial sector. It has been shown that when the simple adaptive expectations model is extended to include more than one expectation variable and when a partial adjustment process is explicitly specified, it is possible to identify the separate contribution of each process. Moreover, since expectation elasticities can be estimated simultaneously with the other structural parameters of the model, it is no longer necessary to impose a priori weights for the expectation generating functions. More generally, the specifying and estimation procedures utilized in this study permitted a relaxation of a number of the restrictive assumptions usually employed in macro investigations of the monetary sector.

Future investigations of the monetary sector will hopefully analyze the consequences of a further relaxation of particular specifying assumptions. One would wish, for example, to modify the assumption that the nominal money supply is exogenously determined, in order to capture the effects of the endogenous responses of the banking system. Since the banking system's response may also depend on expected magnitudes, one could extend the analysis to include a more complete specification of expectation and adjustment processes on the supply side of the market.

This study allowed for different elasticities of expectation for income and the rate of interest; however, it did impose a particular form on the expectation generating equations and on the disturbances in the structural equations. Further investigation of alternative expectation functions and alternative disturbance specifications may yield additional payoffs. More generally, the further application of distributed lag models to the monetary sector is likely not only to improve our powers of prediction but should also enhance our understanding of the structural processes which generate the dynamic responses in the financial sector.

REFERENCES

1. Bronfenbrenner and Mayer, "Liquidity Functions in the American Economy," *Econometrica*, Oct., 1960.
2. K. Brunner and A. Meltzer, "Some Further Investigations of Demand and Supply Functions for Money," *J. of Fin.*, May, 1964.

3. Gregory Chow, "On the Long Run and Short Run Demand for Money," *J.P.E.*, Apr., 1966.
4. F. deLeeuw, "The Demand for Money—Speed of Adjustment, Interest Rates, and Wealth," *Staff Economic Studies* (Bd. of Gov. of the Fed. Res. System, 1965).
5. R. Eisner and R. Strotz, "Determinants of Business Investment," in CMC, *Impacts of Monetary Policy* (Prentice-Hall, Inc., 1963).
6. Milton Friedman, *A Theory of the Consumption Function* (Princeton Univ. Press, 1957).
7. ————, "The Demand for Money: Some Theoretical and Empirical Results," *J.P.E.*, Aug., 1959.
8. Zvi Griliches, "Distributed Lags: A Survey," *Econometrica* (forthcoming).
9. L. M. Koyck, *Distributed Lags and Investment Analysis* (North-Holland Pub. Co., 1954).
10. D. Laidler, "Some Evidence on the Demand for Money," *J.P.E.*, Feb., 1966.
11. Allan Meltzer, "The Demand for Money: The Evidence from the Time Series," *J.P.E.*, June, 1963.
12. J. F. Muth, "Optimal Properties of Exponentially Weighted Forecasts," *J. of the Amer. Statis. Assn.*, 1960.
13. Y. Mundlak, "Aggregation Over Time in Distributed Lag Models," *Int. Econ. Rev.*, 1961.
14. M. Nerlove, *Distributed Lags and Demand Analysis* (U.S.D.A. Agricultural Handbook No. 141, 1958).
15. R. L. Teigen, "Demand and Supply Functions for Money in the United States: Some Structural Estimates," *Econometrica*, Oct., 1964.
16. D. P. Tucker, "Dynamic Income Adjustment to Money Supply Changes," *A.E.R.*, June, 1966.
17. Roger Waud, "Misspecification in the 'Partial Adjustment' and 'Adaptive Expectations' Models" (Univ. of Chicago).
18. A. Zellner, D. S. Huang, and L. C. Chau, "Further Analysis of the Short Run Consumption Function with Emphasis on the Role of Liquid Assets," *Econometrica*, July, 1965.

DISCUSSION

GARY FROMM: Mr. Feige has presented an interesting and provocative paper. This is the first real attempt to distinguish between adaptive expectations and distributed lag adjustments in the demand for money. Unfortunately, Feige has tackled the wrong bird and used inappropriate techniques.

The difficulties begin with the very first equation, that for the equilibrium money stock. Expressed in non-log form:

$$(1) \qquad\qquad M_t^* = \alpha Y_t^{e\beta}.$$

This is a constant elasticity expression that says that equilibrium holdings at the end (or is it the beginning—Feige does not say) of period t is related to expected income in that period. It is not clear (1) why β is necessarily invariant over a long time span; (2) why β is constant over all income levels (Keynesian theory would have it fall as income rises); or (3) why β should be independent of innovations in payments mechanisms, the increasing use of trade credit and industry clearinghouse transfers, the spatial distribution of economic activity, or other factors that affect the velocity of money. While β might be assumed to be relatively fixed over a period as long as a decade, certainly it cannot be assumed to have remained unchanged since 1915.

The second expression, the adjustment of the difference between equilibrium holdings at time t and actual holdings at $t-1$, might also be questioned. Again, in non-log form, desired holdings are:

$$(2) \qquad\qquad M_t^D = M_{t-1} \left[\frac{M_t^*}{M_{t-1}} \right]^{\gamma}$$

Here, the adjustment rate, γ, is independent of the absolute or percentage size of the gap. Thus, there is (1) no recognition of a threshold before action is taken for small gaps, (2) no faster than average adjustment of large gaps, and (3) no distinction between gaps of equal percentage but different absolute magnitude. Also, as above, γ is presumed to be constant over the period of observation.

The third expression, that for the formulation of expectations, suffers from similar constant elasticity difficulties. But, in addition, it has other problems. In non-log form, the expected value of a variable is given by Feige as:

$$(3) \qquad\qquad X_t^e = X_{t-1}^e \left[\frac{X_t}{X_{t-1}^e} \right]^{\lambda}.$$

It is not clear why the expectation in period t, X_t^e, should be based on the realization in that period of actual values to last period's expectation. A more sensible formulation would posit:

474

(4)
$$X^e_{t+1} = X^e_t \left[\frac{X_t}{X^e_t} \right]^\lambda.$$

That is, the expectations for period $t+1$ depend upon the realization of expectations in period t. Then, money holdings at the end of period t for use in $t+1$ can be made a function of expectations in $t+1$; i.e.,

(5)
$$M_t = \phi(X^e_{t+1}).$$

Before turning to the results of implementing Feige's model, it is well to examine the data he employs. For income, he uses Kuznets' annual estimates of real net national product. Because this series includes government expenditures (or on the income side of the accounts, taxes) it is not really an indicator of the volume of private transactions or the level of personal disposable income. Moreover, since desired money holdings presumably are a function of gross cash flows, the exclusion of depreciation is a serious deficiency. As is true of many other studies, interest rates are measured by the commercial paper rate. However, it is dubious whether this is a good proxy for the noncorporate opportunity costs of holding cash balances. And when time deposit holdings are included in the dependent variable, the absence of a time deposit rate term in the equation makes the results even more suspect.

As to the estimates themselves, perhaps the most disturbing of Feige's results is the income elasticity of equilibrium money holdings. For the demand deposits plus currency equation this is 1.3 and for the demand and time deposits plus currency equation, 1.1. These elasticities are three to four times greater than those found by De Leeuw and others. More importantly, they tend to support the quantity theory of money to an embarrassingly significant degree.

Quite frankly, I am skeptical of this result. It would seem that Feige has played the Friedman-Meiselman game in reverse (Commission on Money and Credit, 1963) in a very cleverly conceived and concealed package. Rather than relating real output to the money supply, he makes the money supply a function of real output. Moreover, like Friedman, he uses annual data in level form and fits his equations over a long time span, 1915–63. Thus, the estimates may be nothing more than a reflection of the relative trends (in multiplicative, moving average rate of growth form) in the various series. Certainly, nothing much is indicated about cyclical behavior.

Feige's speed of adjustment estimate (to changes in income) is about one year. De Leeuw's is about $2\frac{1}{2}$ years. Nevertheless, this difference may not be significant since the estimates lie within two standard errors of each other. Yet, there may be a basic difficulty here due to the nature of Feige's model.

The complexion of financial markets in the postwar period is greatly different from what it was in past decades. Not only are there a host of new credit instruments (e.g., certificates of deposit, tax anticipation bills, installment loan paper, etc.), but there are many more financial and nonfinancial intermediaries. Additionally, the impact of discretionary Federal Reserve

monetary policy cannot completely be disregarded, nor can it be assumed that the present-day investor is as unsophisticated as his earlier counterpart.

In this regard, the use of annual observations seems particularly inappropriate. Investment decisions are not made on an annual basis, but continuously. Therefore, the use of monthly, or at least quarterly, data seems advisable. And, if the results are to have any current applicability, only post-accord statistics should be employed. Also, because of the wide divergence in financial behavior of different groups, some disaggregation is desirable. Furthermore, the analysis should be cast in difference form. (Note that subtracting M_{t-1} from both sides of Feige's (19) does not give first difference estimates for the untransformed data. The log of a first difference is not the first difference of a log.)

Moreover, rather than simply positing a single opportunity cost of holding money balances, a more sophisticated portfolio model (along Markowitz lines) might well be utilized. The yields—and their variances—of alternative assets should be included.

Aside from the model and the data, Feige's two-stage estimation techniques may not be sufficiently powerful to distinguish between adaptive expectations and distributed lag adjustments. Given the nature of the decision processes being analyzed, it would seem advisable to check the results using other methods. A Bayesian approach with a priori specification of certain parameters and their distributions might be fruitful. The use of information theory concepts might be even better.

Notwithstanding all the above critical comments, Feige's paper is a good piece of work. His little model is an excellent pedagogical tool and, for the first time, shows the relationship between the adaptive expectations and distributed lag hypotheses.

Professor Christ, too, has presented an interesting paper. He makes the valid point, which few others have stressed, that analyses of government policy actions must consider the interdependence between fiscal and monetary effects. In most other models that include a monetary sector, these interdependencies are implicit. Christ makes them explicit.

However, his model is faulty on several grounds. It is static rather than dynamic. Prices, an important consideration in reviewing fiscal, and especially monetary, impacts, are assumed fixed. Functions are given for the demand for "high-powered" money (whatever that is—currency plus demand deposits?) and government bonds, but all other aspects of the financial structure are ignored.

It is possible to go on listing many more omissions. Yet, even the specifications Christ does set forth can be questioned. Private wealth, w, equals real physical capital plus real high-powered money plus the real value of government bonds. But, surely, only the physical capital is truly wealth—the rest are just paper claims on that capital (including plant, equipment, housing, inventories, land, gold, etc.). Nevertheless, let us grant that some such "perceived" wealth concept might reasonably appear in a consumption function. Why then should the yields on perpetual government bonds and on physical capital appear in the latter in simple, linear fashion?

Leaving the structure of the model and the "guesstimates" about mar-

ginal coefficients aside, one must also doubt the validity of the numerical results given Christ's assumptions on initial conditions. These first appeared in a longer version of the present paper written by Christ in November, 1965. Thus, his assumptions might be taken as estimates of actual figures in 1965. However, for some variables the divergences are so wide that it would be unwise to take the results as being applicable to that year, if any other. For example, the difference in real NNP (actual minus Christ) was $17 billion; in government expenditures, $15 billion; in government bond rates, 90 basis points; and in disposable personal income, minus $15 billion. Personal consumption expenditures, on the other hand, coincided almost exactly with $429 billion for the actual and $430 billion for the Christ figure.

But, let us even forget about these difficulties and ask about the meaning of Christ's "multipliers." He divides changes in real NNP by changes in exogenous policy flows and treats these as if they were traditional multipliers. First, it must be observed that real GNP rather than NNP normally is used in the numerator. Second, the denominator generally is an exogenous increment, decrement, or net transfer of real resources. Christ's denominators, on the other hand, are not wholly exogenous, for he sometimes credits his policies with induced reductions in government debt. Also, he treats open market purchases and the printing of high-powered money as if they were equivalent, dollar for dollar, to real resource increments. Why the private sector should view an exchange of bonds for money as an increase in its resources is not clear. Assuredly, such a switch has implications for interest rates and time and demand deposit holdings. And, these can be translated into equivalent real inputs. (Methods for doing so are contained in my own paper for these meetings.) With such equivalents inputs, it is then possible to compute multipliers that are comparable to those usually examined by the profession.

For all its shortcomings, Christ's paper is still a valuable contribution to the literature. However, rather than making up his own model and fanciful estimates, Christ would have been better advised to utilize one with firmer empirical foundations (e.g., Friend-Taubman, Liu, or OBE in the small model range), modifying it to reflect the desired fiscal-monetary constraints.

THOMAS R. SAVING: All three of these papers concern themselves with the effectiveness of countercyclical policy, especially monetary policy. However, effectiveness is not defined in the same way in all three papers. In Christ's paper the method of analysis used is comparative statics and thus effectiveness is defined as the magnitude of the policy multiplier. In the papers by Feige and Shapiro, effectiveness is defined as the speed of adjustment to policy, and thus these papers are, at least in this respect, dynamic. The score on effective monetary policy from this sample of three is two for, one against. Let me discuss the papers in alphabetical order by author.

I found Christ's paper very interesting. It is pleasantly short and to the point, which is that policy, especially fiscal policy, cannot be analyzed without taking account of the fact that the government has a budget constraint. Christ correctly points out that this constraint is neglected in most macroeconomic analyses. Moreover, as I am sure Christ realizes but fails to mention, the

budget constraints of the public and business sectors are also left out of most analyses. Thus, it seems that in spite of the general impact that Patinkin's *Money, Interest and Prices* has had on the profession, the constraint imposed by Walras' law has not been generally incorporated into macroeconomic analyses. However, a modified version of Christ's budget constraint has been used to analyze monetary policy by Mundell in a 1960 paper (*J.P.E.*, Dec., 1960) and Pesek and myself in a 1963 paper (*J.P.E.*, Aug., 1963). Additionally, as Pesek and I show in a manuscript to be published in January, 1967 (*Money, Wealth and Economic Theory*, Macmillian, 1967), the failure to consider this constraint accounts for the entire difference between Gurley and Shaw's "inside" and "outside" money. However, beyond the inclusion of the government's budget constraint (and this is an important contribution), Christ's paper does not add materially to the discussion of the effectiveness of monetary policy, since it simply confirms the generally accepted result that in a static equilibrium model, the money multiplier is nonzero and may be large. The current controversey concerns the dynamics of the system rather than the statics, with emphasis placed on the speed of adjustment to changes in money market variables. Essentially, the point being the question of whether or not monetary policy affects aggregate demand rapidly enough to counteract, for example, a decline in aggregate demand. Or will this policy take affect with a long lag that may result in feeding a later boom? This is the point to which the Feige and Shapiro papers address themselves.

Feige's paper is interesting in that the implied lag in the behavioral adjustment of consumers and firms is very small. In fact, his results indicate that the entire adjustment of actual cash balances to desired cash balances takes place in a single period. In the paper Feige notes the distinction between the treatment of money as a capital good that is held for the services it yields versus the treatment of the traditional transactions motive. However, the two approaches are identical except that the service that money yields is restricted to services as a medium of exchange in the transactions case unless one assumes that the transactions time is zero. In this case, money will not exist. Money by its nature must be held during transactions and because of the cost of converting other assets into money for every transaction money must be held before transactions. But then money is being held for the services it renders, and this is the capital theoretic approach that Feige uses. In both the formulation and statistical work Feige has done a careful and well-thought-out job. His consideration of the costs of disequilibrium is a very interesting one but somewhat incomplete. The consistency of his estimates of the lag structure through all the estimating procedures lends some credence to his results.

The third paper, presented by Harold Shapiro, has results remarkably different from Feige's. The difference is remarkable because of the similarity between the approaches used by the two pieces of work. The work by Feige employed a more complete model but the actual statistical work was similar. Shapiro's work was done in terms of ratios of variables (velocity) rather than the variables themselves so that some statistical questions arise in comparing his results with Feige's. Also, his equation (1) contains an error which fortunately was not carried forward into the remainder of the work.

The estimate of the proportion of the difference between actual and desired

cash balances that will be eliminated in a single period is approximately 0.4 in Shapiro's work or approximately 0.65 on an annual basis, while Feige obtained estimates of the same proportion that were close to one, indicating, as pointed out earlier, that the entire adjustment would occur in a single period. While these two studies did utilize different data (Feige used annual U.S. data and Shapiro used quarterly Canadian data), this is a remarkable difference. It may be that had Shapiro considered a more complete model with a supply side (this may be especially important in an open economy such as Canada) the difference in the results may have disappeared.

Shapiro also makes the argument that monetary policy primarily affects the short-term rate of interest and that the fact is that this is not the rate of interest that enters aggregate demand relations. He offers no proof of this assertion other than a statement that business investment decisions are thought to be function of present rates and expected future rates, which implies that the long rate only is relevant. However, this would not be implied by a theory based on firm behavior designed to maximize discounted future profit streams. Such a theory would include both the short and long rates, so that the timing of investment decisions may be altered by changes in the short rate only. However, Shapiro correctly maintains that if long rates are affected very little by monetary policy, the effect on investment of the short rate changes that are induced by policy will be smaller. He tests the effect of the short-term rates on expectations of future rates, and thus the long rate, by making use of the preferred habitat theory. He concludes that while there is a considerable "extrapolative" element in the relationship between the current short rate and expected future rates, this is less true than it was in the period before the accord. However, his results still show a substantial relationship between the short and the long rate.

GEORGE R. MORRISON: The principal contribution—and an important one it is—of Professor Feige's paper is to provide an unambiguous statistical discrimination between expectations lags and adjustment lags in the demand for money. Confounding of the two lags has plagued previous studies of demand for money, and, indeed, most other recent attempts to fit stock demand equations of any sort. Feige makes use of a nonlinear least squares regression routine to identify the structural expectations and adjustment parameters that remain overidentified when the standard Koyck-Nerlove stock adjustment-cum-adaptive expectations model is fitted by ordinary least squares techniques. The essential point is to recognize that in this context, minimizing the sum of squared residuals from the regression surface must be done subject to certain nonlinear constraints that are imposed on the regression coefficients of the reduced form equation in such a manner as to force them to yield identical solution values for the structural parameters, no matter how the latter are derived from the regression coefficients.

There are other respects in which Feige has utilized fairly rigorous econometric techniques in the hope of avoiding the pitfalls of statistical inference in dynamic economic models. To avoid simultaneous equations bias, the demand for money equation is first incorporated in an income-expenditure framework in which the supply of nominal money is specified to be a function of autono-

mously determined high-powered money and the discount rate, and aggregate expenditures are a function, in part, of autonomously determined government expenditures. Then the demand equation is estimated by two-stage least squares, using the autonomous items as instrumental variables. To guard against drawing erroneous inferences from regressions in which the residuals are serially correlated, Feige tests for and rejects the existence of first order autocorrelation in the reduced form residuals.

Nevertheless, I cannot accept these procedures without some reservations. What Feige ends up with is a regression in which there are 10 independent variables (including four lagged values of money stock, three values of income, two of them lagged, and three values of the interest rate, two of them lagged) and 49 annual observations on some highly trend-dominated series. One wonders whether such a regression practically guarantees that the residuals will be negligible and the serial correlation undetectable. How many degrees of freedom can be squeezed from these time series turnips? In interpreting the results of models such as Feige's, it is well to bear in mind Griliches' dictum that "in most cases the addition of the lagged dependent variable to the regression will reduce the serial correlation of the residuals. But unless one assumes that the serial correlation is due to the distributed lag effect, this is no evidence for the distributed lag model. There may be other reasons for serial correlation."

Turning to matters of economic substance, Feige concludes, from some preliminary spadework, that the demand for real cash balances is best expressed as being homogeneous of degree zero in population and/or the general price level. The demand for real cash balances (either M_1 or M_2) is then found to display a permanent real income elasticity of close to unity, and an elasticity of less than -0.2 with respect to the expected short-term interest rate. The expectations lags for income turn out to range between thirty-one months (for M_1) and forty months (for M_2), while for interest rates it is from fourteen to sixteen months. Feige's most noteworthy result, and one that differs from some previous findings, is that the adjustment lag is probably no longer than a year for either M_1 or M_2.

These are not unreasonable results. If I were to quarrel with any of them, it would be with the estimate of the adjustment lag. My own findings, which (as explained below) are based on a somewhat different model of short-run response, would point to a lag of at least half again as long. But the difference in results could easily be due to differences in specification, so I am not inclined to be dogmatic on this point. I might also point out that the deflation of cash balances and money income by the usual implicit price index, as Feige and most others have done, conceals the fact that virtually the same results will be obtained if the deflation is not performed and, indeed, if the price index is not even included in the regression. This leaves the economic interpretation of the observed income elasticity somewhat in limbo, as a fortuitous mixture of real and nominal effects on the demand for money. In my opinion, the problem of the proper deflator for money calls for considerably greater attention by empirical workers than it has hitherto been accorded. A somewhat lesser quibble has to do with the measurement of expected returns on holding

assets other than money. Instead of calculating the distributed lag from cur-
rent and past values of the short-term interest rate, it might be better to as-
sume that the expected return is composed of a weighted sum of the current
bond yield and an expected rate of change in the price of bonds. Feige's meth-
od admittedly has the advantage, in his model, of avoiding the further prolif-
eration of lagged variables.

When it comes to the basic economic meaning of his results, Professor Feige
presents some of the aspects of a moving target. In the first draft of his study
he argued that because he found the demand for money to be a function of
permanent rather than current measured income, his results confirmed the
view that the durable goods or asset motive dominated over the buffer stock or
transactions motive for holding money. Now, he has abandoned this conclu-
sion, in favor of the view that permanent income may simply be a better esti-
mate (optimal forecast) of income rather than a proxy for wealth. Neverthe-
less, Feige's sentiments clearly lie with the view that money is a durable good,
not a buffer stock. Drawing support from a quadratic cost minimization model
of the demand for money, he suggests that speedy adjustment of actual to de-
sired cash balances is evidence for the relative unimportance of transactions
costs in the demand for money.

Yet Feige's view is not convincing. In defense of the buffer stock hypothe-
sis, there are several things to be said. In the first place, money may be held
for both durable asset and buffer stock reasons. Second, it is not generally un-
derstood that a buffer stock hypothesis requires only that transitory distur-
bances of any kind—be they transitory income, interest rate, or price level
movements, or transitory movements in the money stock itself—find reflection
in desired cash balances.

I have found strong empirical support for a buffer stock hypothesis which
postulates that holders of money react passively to transitory disturbances to
the nominal stock of money by letting their actual cash balances depart from
the long-run relation to permanent income, prices, and costs of holding money.
To attempt to maintain the long-run relation in the face of these transitory
money disturbances would impose costs, both pecuniary and psychic, of rapid
alterations and reversals of expenditures and portfolio distributions. In testing
this hypothesis, I find, like Feige, that Friedman underestimated, but only
slightly, the lag in expectations for permanent income that is relevant to the
demand for money, and also like Feige, I find the long-run income elasticity of
demand for money is closer to unity than to the 1.8 estimate of Friedman
(about 1.3 according to my estimates). But I also find that to obtain these re-
sults, it is necessary to introduce transitory money into the demand equation,
with an implied lag of eighteen months for permanent money, or roughly
half the lag for permanent income.

Time does not permit discussion of certain interesting properties of transito-
ry money as a link between studies of the long-run demand for money and dy-
namic monetary explanations of business fluctuation, especially those that
suggest the existence of a lag between turning points in money and turning
points in business.

THE MEASUREMENT OF PRICE CHANGE
A REPORT ON THE STUDY OF INTERNATIONAL PRICE COMPETITIVENESS*

By IRVING B. KRAVIS, *University of Pennsylvania*
and ROBERT E. LIPSEY, *National Bureau of Economic Research*

I. *Aims and Methods*

This paper is a preliminary report on the overall results of the National Bureau's study of comparative prices and price trends in the international trade of the United States and its main foreign competitors. A discussion of the aims and methods of the study, along with tentative results for iron and steel products, was published in 1965, and a report on nonferrous metals has just been published by the National Bureau.[1]

The purpose of the study has been to develop methods of measuring price competitiveness in the international trade of a developed industrial country. These methods have been applied to trade in machinery, transport equipment, and metals and metal products for the period 1953-64.

There is a need for better measures of price competitiveness because we now have no measures of relative price levels and because existing measures of price changes in world trade, such as unit value and wholesale price indexes, have serious deficiencies which we discussed in our earlier papers.

The deficiencies of the existing indexes suggest three specifications to make a price index appropriate to the study of international competition: (1) it should be based on actual prices or price offers, not on list prices or unit values; (2) for goods which the country exports, the prices should refer to export rather than domestic transactions; (3) the universe of prices should include prices of all goods, in the classes under study, that enter world trade. To fulfill this requirement, domestic prices are taken for products which a particular country does not

[1] Irving B. Kravis, Robert E. Lipsey, and Philip J. Bourque, *Measuring International Price Competitiveness: A Preliminary Report*, Occasional Paper 94, 1965; Irving B. Kravis and Robert E. Lipsey, *Comparative Prices of Nonferrous Metals in International Trade, 1953-64*, Occasional Paper 98, 1966.

export, and the weighting system is based on the relative importance of commodities in international trade.

The new measures are made up of three interrelated sets of index numbers:

1. *International Price Indexes.* These are time-to-time indexes for each country, derived by applying 1963 world trade weights to each country's export prices. They measure the change in each country's prices of the bundle of goods that was exported by the industrial countries as a whole.

2. *The Index of Price Competitiveness.* These are comparisons of price movements, made by dividing the international price index for each foreign country by the corresponding index for the U.S. We call the resulting index an index of U.S. price competitiveness relative to that country. A rise in an index of U.S. price competitiveness, therefore, indicates that the foreign prices have risen relative to U.S. prices; U.S. price competitiveness has thus improved while that of the foreign country has declined.

3. *Comparisons of Price Levels.* The index of price competitiveness can also be derived from country-to-country comparisons of price levels at a given moment in time. In order to compute the index of price competitiveness from the place-to-place price relatives, the ratio of foreign to U.S. prices for each year is taken as a percentage of the ratio for the base year.

An important technical feature of the study was the decision to abandon the usual practice of organizing the price collection effort around a set of product specifications selected in advance. We placed the burden of determining comparability on the respondent, asking him to select the most important items and to provide comparable quotations either over time or between exporting countries. We ourselves rarely undertook to match two prices in order to compute a time-to-time or place-to-place price relative. In a sense it was the price relative itself that we were collecting in our field work, although we did obtain the actual prices as well.

The comparative prices used in this study were gathered from a variety of sources.

A major portion of the data came from more than 200 American firms, mainly large companies. U.S. sellers of machinery and metal products were asked to provide their own export prices and to compare them with those charged by foreign subsidiaries, licensees, or competitors for identical or equivalent products. Companies involved in international markets through their purchasing activity were asked to compare offers from the U.S. and foreign countries for specific items of

equipment or metals and to trace the changes in the prices of such items over a period of years.

Price data were also obtained from a number of governmental agencies. These data, which consisted largely of formal bids by U.S. and foreign firms to supply metals or equipment, were collected, with a great deal of help from the Bureau of Labor Statistics, from government-owned utilities, the military services, and other federal agencies.

The third major body of data was from foreign sources. The most comprehensive time-to-time data, from Germany, consisted of the official export and import price series broken down in considerable detail. The German data included a limited number of direct place-to-place price comparisons from private sources. Other collections of comparative price information were obtained from Israel, Thailand, and the United Kingdom.

We were also able to obtain additional information on bidding for contracts to supply foreign agencies, mostly governmental, with a wide variety of machinery and equipment. More than 1,300 such bids, some of which ran into millions of dollars and provided international price comparisons for many items, were analyzed.

We believe that these varied sources of price data provided good samples of the basic metals, metal products and machinery included in our study. For some important commodity segments the sample covers a substantial fraction of international trade. All in all at least some data are included for purchases of firms or public agencies in each of about fifty countries.

Much of our confidence in the results of the study rests not only on the large number of observations but also on the variety of sources. Data from each individual type of source may be subject to biases of unknown importance, but there was a good chance of overcoming most of these by including a large number of almost every type of transaction that involves world trade.

II. *A Sampling of Results*

We plan to produce, by the time the study is complete, price indexes showing changes over time and country-to-country differences in price levels for most of the three-digit SITC[2] groups included in the study and for many of the more important four-digit subgroups as well. So far, we have tentative indexes for four SITC divisions which include about two-thirds of the value of trade covered in the study, and these should be sufficient to give some notion of the way the final results will come out. For these four main product groups, we summarize our

[2] United Nations, *Standard International Trade Classification, Revised,* Statistical Papers, Series M, No. 34 (New York, 1961).

findings on the comparative level of prices of internationally traded goods in 1964, the final year of the study, and the movement since 1953 of the indexes of U.S. price competitiveness.

In 1964, international price levels of the United Kingdom and the EEC countries were below those of the U.S. for every one of the four groups. The U.S. price position was most unfavorable for iron and steel products, with foreign prices about 20 percent lower, and best in nonferrous metals, for which the price differences were very small. For the two groups of advanced manufactures the differentials were all less than 10 percent: 8 to 9 percent for transport equipment (other than road motor vehicles), and 5 to 8 percent for nonelectrical machinery.

The summary indexes, shown in Table 1, conceal a good deal of

TABLE 1

COMPARATIVE INTERNATIONAL PRICE LEVELS FOR FOUR MAIN PRODUCT GROUPS, 1964*
(U.S. = 100)

SITC	Commodity Division	Country			
		U.S.	U.K.	EEC	
				Total	Germany
67...............	Iron and steel	100	81	80	81
68...............	Nonferrous metals	100	98	98	100
71...............	Machinery other than electric	100	92	96	95
73 (excluding 732).	Transport equipment (excluding road motor vehicles)	100	91	92	91

* These indexes are averages of indexes for smaller groups on the three-, four-, or five-digit SITC level, weighted by the importance of each subgroup in world trade. The price level indexes for SITC 67, 68, and 73 other than 732 include estimates for almost all the three-digit components included. They may still be subject to some revision. The index for SITC 71 is composed of data for subgroups including only about two-thirds of trade for the group, and we hope to raise this proportion considerably in our final report.

variability in subgroup price ratios and, of course, more in the ratios for individual products. However, the ratios do cluster around the averages shown, with the greatest number of subgroups of iron and steel, for example, and those with the greatest value of trade, showing U.K. and EEC price ratios between 80 and 85 percent of U.S. prices. Japanese indexes for iron and steel, not shown in Table 1, point to a price level considerably below the European one, at about 25-30 percent beneath the U.S. price level.

In the case of nonferrous metals, the price relationships for the two largest subgroups, copper and aluminum, were fairly favorable to the U.S. in 1964, and foreign prices of lead, zinc, and silver were also close to the U.S. level, or, in some cases, above it. However, nonferrous met-

als prices are subject to wide fluctuations and the price relationships in 1964 were more favorable to the U.S. than those of most previous years.

Both nonelectrical machinery and transport equipment contain a wide range of price level ratios. European prices were higher than those of the U.S. in 1964 in five out of eighteen subgroups for the U.K. and eight out of eighteen for the EEC countries.

The widest contrasts are within SITC Division 73, in which prices offered by European and Japanese builders of ships and boats were around 50 percent lower than U.S. prices, a difference considerably greater than that for the worst of the iron and steel groups in the years in which the U.S. competitive position was most unfavorable. This same SITC division also contains, on the other hand, railway vehicles and aircraft, two groups in which the price competitiveness of the United States was high.

Changes in Price Competitiveness. The largest changes in international price competitiveness, that is, the relationship of foreign prices to U.S. prices of internationally traded goods, have taken place in iron and steel. Mostly these have been unfavorable to the United States, with a decline of almost 20 percent in U.S. price competitiveness relative to the European countries between 1953 and 1963 (Table 2). In 1964, however, there was a reversal of the trend, with U.S. price competitiveness rebounding to better levels than in any of the three preceding years. Relative to Japan there was a decline in U.S. price competitiveness between 1961 and 1962 and no change after that. Despite the gain in 1964, however, the position of the United States in iron and steel remained considerably worse than in 1953 and 1957.

Nonferrous metals, too, saw a decline in the price competitiveness of the United States relative to the EEC between 1953 and 1963, but there was only a little gain or no significant change relative to the U.K. However, the fall in EEC relative prices was much milder than for iron and steel, and the price competitiveness of the U.S. appeared better in 1964 than in any previous year.

Nonelectrical machinery, which has a heavier weight in our total than the other three groups combined, showed only small changes in price competitiveness. The U.S. relationship to the U.K. can be read as showing no change at all over the period. Relative to Germany and the EEC as a whole there was a deterioration between 1953 and 1957, and then a recovery, with very little change after 1961.

Only in the transport equipment group were there substantial differences between U.S. price competitiveness relative to the U.K. and that relative to the EEC. After an initial decline U.S. price competitiveness relative to the U.K. rose by more than 15 percent to 1964.

TABLE 2

INDEXES OF U.S. PRICE COMPETITIVENESS* RELATIVE TO THE
U.K., GERMANY, AND THE EEC, 1953–64
(1962 = 100)

	1953	1957	1961	1962	1963	1964
SITC 67 Iron and Steel						
Relative to U.K.............	120	110	101	100	97	103
Relative to EEC............	121	119	103	100	98	105
Relative to Germany........	115	112	103	100	96	104
Relative to Japan...........	n.a.	n.a.	108	100	100	100
SITC 68 Nonferrous Metals						
Relative to U.K.............	100	101	100	100	102	107
Relative to EEC............	105	102	100	100	101	109
Relative to Germany........	106	105	100	100	99	108
SITC 71 Nonelectrical Machinery						
Relative to U.K.............	99	101	100	100	100	100
Relative to EEC............	102	97	98	100	99	99
Relative to Germany........	100	95	98	100	99	99
SITC 73 (exc. 732) Transport Equipment, except Road Motor Vehicles						
Relative to U.K.............	96	91	101	100	105	106
Relative to EEC............	106	96	97	100	98	97
Relative to Germany........	104	101	96	100	98	97

* These indexes show the relationship between the change in foreign prices and the change in U.S. prices. For example, the figure of 103 for SITC 67 relative to the U.K. in 1964 shows that U.K. prices rose 3 percent more than U.S. prices since the 1962 base year, or, if derived from price level data, that the ratio of the U.K. price level to the U.S. price level was 3 percent higher in 1964 than in the base year, 1962.

Indexes for SITC 67 represent a revision of those published in *Measuring International Price Competitiveness*, mainly the result of adding large amounts of new data, especially for 1963 and 1964. SITC 68 indexes are from *Comparative Prices of Nonferrous Metals in International Trade, 1953–1964*. Indexes for all four divisions are calculated mainly from time-to-time data for three- and four-digit SITC subgroups with some use of place-to-place data. Subgroups included in the indexes generally account for at least two-thirds of the value of total trade in each division; the final estimates will improve the coverage, especially for SITC 71.

Relative to the EEC countries, however, U.S. price competitiveness never recovered from its early fall and ended the period showing a loss of over 5 percent.

On the whole, the larger movements of an index of price competitiveness appear to represent not only average changes for a group but also a substantial degree of consensus among the subgroups. In iron and steel, for example, the proportion of subgroup indexes of price competitiveness rising or falling was closely related to the movements of the group indexes for both the U.K. and the EEC countries.

It would seem reasonable to expect much greater diversity in indexes of price competitiveness among machinery groups than among subgroups of iron and steel, since the different kinds of machinery are in many cases produced by quite separate industries supplying very

different markets. In fact, the range of the indexes of price competitiveness in SITC group 71 is quite narrow in most years. Of sixty-five year-to-year changes in U.S. price competitiveness relative to the U.K. and EEC for three-digit and four-digit components of the machinery group, more than half were one percentage point or less. In only eight cases were there changes in price competitiveness by more than three percentage points. The competing exporters' prices in these groups apparently moved very closely in unison, to a much greater degree than in metals.

The proportion of machinery groups showing improving U.S price competitiveness in each year is a more volatile index than the average changes themselves. Its movements resemble those of the corresponding proportions for subgroups of iron and steel.

Taking the indexes of Table 2 as a group, we can consider only 1963-64 as a period of predominantly rising U.S. price competitiveness. This impression is reinforced by the unanimity among the subgroups regarding the direction of change. The rest of the period was characterized by declines in the price competitiveness of the U.S., especially with respect to the EEC countries. The major declines were between 1953 and 1961, and were followed by comparative stability (except in iron and steel, where the U.S. position continued to deteriorate) and a shift toward a mixture of rises and falls before the gain in 1964. That improvement, visible in the indexes for metal products, showed up in machinery and transport equipment only as a gain in the number of subgroups reporting an improvement in the U.S. position.

III. *Some Tentative Conclusions*

A major conclusion of our work at this point is that it is feasible to collect a great variety of data relevant to the measurement of international price competitiveness that have never been collected before. Working with comparatively small price collection resources we have accumulated a large quantity of data. A well-financed official data gathering project should meet with even greater success.

The indexes of relative levels of and changes in price competitiveness we have calculated from these data often move quite differently from unit value and domestic wholesale price indexes. On logical grounds our indexes have a strong claim to be superior measures of competitive strength, and in some cases we have evidence that specific unit value series, in particular, give a false picture of price changes. Whether the new measures will turn out to have great predictive or explanatory value is still to be determined by attempts to use them in the analysis of trade structure and trade flows.

It seems clear that future efforts at data collection in this field

should be conducted in more than one country and preferably in many countries. Comparisons in any one market tend to be biased by trade barriers, consumer preferences, differences in the degree of competition and in the range of products purchased, and many other factors. The ideal arrangement would be for an international agency to act as a clearinghouse through which countries could exchange data derived from government purchasing activity and from firms operating in the individual countries, with each government and firm reporting on its foreign as well as its domestic operations. It would be important to collect data from the less developed countries, since these are major markets for many products. However, even an exchange between any two or three countries could add greatly to the information available to each one on its competitive position.

For both producer and consumer durable goods we have found that the use of regression methods for international price comparisons permits the coverage of many complex products which would defy comparison by conventional methods involving identical specifications in two situations. For many such goods there are no cases of identical specifications from two countries or, in some products, from two periods of time. We have applied regression methods to such products as locomotives, aircraft engines, automotive diesel engines, outboard motors, tractors, chemical reactors, automobiles, and ships, most of which would have presented insurmountable difficulties for comparison via identical specifications. The method essentially involves treating a commodity as a cluster of characteristics or quality elements and measuring the price of each characteristic through multiple regression of price against the amount of each element. We have applied it to numbers of models ranging from 20 to 1,000, using 1 to 6 physical characteristics of the product (but often experimenting with more) and have usually reached $\bar{R}^2$s of 0.9 or better in the best equations. Although we have used regression methods mainly for international comparisons, the fact that we were able to obtain satisfactory results with such limited data augurs well for the use of these methods in domestic price index work as well. The application of such methods might help to eliminate an important source of potential upward bias in the present indexes.

Another experiment we consider successful and more broadly applicable was the abandonment of the traditional method of using preselected specifications and the placing of the burden of finding comparable products on the respondent. Especially for more complex products it would have been impossible to pick specifications applicable to any large number of respondents. The use of bidding data in a sense followed the same method, since each bid comparison was for a very par-

ticular product or set of products, and it was rare that any two were for items that were identical to the degree required by present price collection methods. No set of preselected specifications could have turned up more than a small fraction of the items appearing in these bids.

In the course of examining the great number of different bidding and selling arrangements, we were frequently reminded of the enormous variety of conditions attached to a sale, other than those usually subsumed under the heading of "price." List prices were, of course, subject to many types of discounting, for cash payment, for size of order, or simply to meet expected competition. Often this discounting was ignored in sellers' reports and only revealed in buyers' reports and in bidding documents, and the prevalence of this discrepancy points strongly to the need for collecting data from buyers as well as from sellers. Even the transaction price, however, should be regarded only as a reference base from which continual additions and subtractions are made through changes in such factors as credit terms, delivery time, and the provision of various services. All these features could conceivably be priced, but the information is difficult to obtain, and the lack of it probably causes us to underestimate the real degree of price flexibility.

Another variable which seems impossible to measure, but plays a role in the success of American exports, is the degree of confidence in the quality of American products established by technological leadership. Purchasers of a number of products reported to us that U.S. products (bearings and pumps, for example) were sometimes purchased, despite their higher prices, when critical uses were involved. The factor involved here was not only the average level of quality, but the confidence of the buyer in the consistency of the quality, or in the small risk of failure to meet the required standard.

In many cases speed of delivery has been an important advantage for American firms. We have made a substantial collection of delivery data and hope to include an analysis of it in our final report. It does appear that this is an element in competitiveness and that it is not only a cyclical phenomenon, but one which persisted, although with fluctuating magnitude, over the whole period covered by our study. This factor appears sufficiently important to warrant inclusion in future examinations of international competition.

The examination of price data, particularly country-to-country comparisons for specific products, reveals the inadequacy of the available trade classifications for analyzing trade in advanced manufactures. The United States tends to be strong in technologically advanced products, while products whose technology is well established are often

cheaper abroad. Many of these differences, however, operate within the available trade classifications, even the five-digit classifications of the SITC. As a result, the trade statistics give the appearance of a large degree of cross-exporting in many commodity groups. This is true particularly in the cases where the greater technological sophistication of American industry makes it possible both for technological advances to occur early in a given industry and for the industry to find customers who are ready to use a more sophisticated product. In some such cases American companies producing at home and abroad recognize these differences by systematically lagging the production of new products abroad by one "product generation." We have found this to be true, for example, in the case of office machinery and construction machinery. A comparison involving only the older item, which is the only one produced in both countries, gives a misleading impression of the competitive position of the United States.

It appears likely at this point that the strength of the U.S. trade position in machinery and related products rests mainly on the availability in the U.S. of products more sophisticated or technologically advanced than those produced abroad. The United States lead is accounted for in some industries by a lead in knowledge and in others by the scale of the American economy which makes the production of larger and more efficient machinery economical in this country before it is feasible abroad. There are often no exact foreign counterparts to these U.S. products to compare by conventional methods, and the closest competing products, when account is taken of efficiency through regression methods or buyers' analyses of bids, frequently prove to be more expensive than the U.S. products.

We hope that the outcome of this investigation will encourage government and international agencies to pursue the measurement of international price relations on a more comprehensive basis. Such measurements would add to our understanding of trade patterns and of changes in the balance of payments of industrial countries and should also be useful in analyzing the shifts in trade for specific groups of commodities. Our experience has persuaded us that the job can be done and that indexes like these should prove to be a great improvement on existing measures.

THE CONSTRUCTION OF INDUSTRIAL PRICE INDICES*

By JAMES K. KINDAHL

University of Chicago

I

Observers have long suspected that some of individual price series in the Wholesale Price Index may not be the prices at which goods actually move in volume sales. The Price Statistics Review Committee in its well-known *Report*[1] of 1961 cited fragmentary evidence which suggested that divergences between reported prices and actual prices may be of substantial magnitude. That markets exist in which such a divergence occurs is beyond doubt. These divergences are due partly to the reluctance of sellers to disclose the facts, partly to the fact that a catalogue price is usually obtainable without difficulty whereas the true selling price, including all discounts, overt and covert, may require some figuring and searching to obtain and may differ from customer to customer or transaction to transaction. One way to circumvent the difficulty is to obtain prices from a sample of buyers of commodities rather than exclusively from sellers as has been done for the WPI in the past. This procedure was strongly recommended by the Price Statistics Review Committee. The National Bureau of Economic Research Study of Industrial Prices, being conducted by George Stigler and myself, is based on this approach. (Recently, the BLS announced that it was studying the index to determine commodity areas in which this approach might be desirable and feasible.)

There are a number of differences between indices based on sellers' reports and those on buyers' reports. Four important ones are discussed here.

1. Even when freight rates remain constant for given shipments between given points, sellers' receipts (mill net receipts) may move differently from buyers' payments. Many industrial goods are quoted as delivered prices, or as f.o.b. mill with freight equalization. In both cases, mill net receipts for a given invoice payment vary with the destination of the goods. In the short run at least, average distances of shipment can and do vary even though delivered price does not. Thus, for example, each firm in the steel industry quotes prices f.o.b. its own mill but equalizes freight with the mill of any firm nearer to the cus-

[1] N.B.E.R., Price Statistics Review Committee, *The Price Statistics of the Federal Government*, N.B.E.R. General Series No. 73 (1961).

omer—which is to say, prices are in effect quoted f.o.b. the mill near-
est the customer. In periods of high operation, steel firms tend to con-
centrate their sales efforts near their own mills; in periods of weak de-
mand, they tend to range further afield. Mill net prices show a definite
cyclical pattern, declining in contractions and increasing in expansions.
Published f.o.b. mill prices show no such pattern. This variation can be
due only to covert price cutting and to increased freight absorption. At
present there are no published data which allow estimates of the rela-
tive components of the two. It is expected that the N.B.E.R. study will
shed light on this question.

Thus one might record three prices for each transaction: the net
price (mill net) received by the seller; the delivered price paid by the
buyer; and the (sometimes only theoretical) f.o.b. mill price. The
f.o.b. mill price seems to be the one favored by many analysts, as
being a "pure" price not subject to the vagaries of freight costs. To the
extent that it differs from the other two, however, it differs from the
price as seen by either party to the transaction—a fault which would
seem to be disqualifying.

An index of industrial prices could be used to measure the costs of
inputs into industrial processes; such an index would be appropriate as
a current indicator of cost changes and as a meaningful basis for con-
tract price adjustments deflated by the WPI or its components.
Similarly an index of prices received by sellers would be appropriate to
use for deflation of value of output. It is not clear just what purposes
the theoretical f.o.b. mill price can serve in cases when it is different
from the mill net receipt, except as a compromise between the two
other measures, or as a figure for which data may be relatively easy to
obtain, which may be used as an approximation to one of the others.

2. In sampling buyers rather than sellers for price information, it
becomes practically impossible to use the extremely detailed
specifications prescribed by the BLS for the present WPI. In order to
find a substantial number of customers who meet the requirements—
not only for particular models or grades, but particular sizes of the
goods and particular transaction lot sizes—a huge number of purchas-
ers would need to be canvassed. It is necessary to use broader clas-
sifications; e.g., hot rolled carbon steel sheets rather than "hot rolled
carbon steel sheets, 10ga. × 48" wide × 120" long, sheared edge, cut
length, base chemistry, commercial quality, base packaging, base
quantity, mill to user, f.o.b. mill," as the WPI does. Of course, any one
reporter will have to be held to tight specifications, to avoid confusing
a change in quality of goods with a true price change.

The present BLS procedure, from a sampling standpoint, is essen-
tially as follows. Out of the wide variety of specific items which go to

make up a class of commodities (e.g., aluminum mill shapes), one or more specific products are chosen to represent the class. For each of these products, precisely specified, price quotations are obtained from one or more sellers. Thus a sample of specific commodities is drawn from the commodity class. If n reports are to be collected in total, obvious alternatives to the BLS practices are to obtain one price on each of n commodities or to obtain a smaller number of reports on each of a larger number of commodities. So long as the commodities chosen remain constant over time, either procedure is compatible with a fixed base index.

Sampling considerations taken alone suggest that the choice be made so as to achieve minimum variance of the estimates thus obtained. There is no obvious reason why a more extensive sampling of commodities, with a correspondingly less intensive replication of prices of each commodity, is inferior to the BLS procedure from this point of view. Thus the departures from BLS sampling procedures which are required by a buyers' price approach are not necessarily undesirable changes.

3. It not infrequently happens that a firm manages to secure a reduction in the price of a given commodity by shifting from one supplier to another. Such a change is indeed a reduction in price to the buying firm and should be counted as such. If the old supplier continues to quote the higher price to other customers, a fixed weight index of sellers' prices will not show a price decline, whereas a fixed weight index of buyers' prices will. (The discrepancy arises from the use of fixed weight; an index with current period weights for each reporter would show the same results for both a sellers' and a buyers' price index, because of the shift in weight involved.) In this case, the buyers' price index procedure is surely superior to the sellers' index procedure.

4. One fact brought out clearly by the N.B.E.R. study is the widespread use of "purchase contracts" in industrial purchasing. These contracts are for varying periods of time—usually three, six, or twelve months (sometimes longer) with twelve being the most common—and obligating the seller to supply a given amount of a commodity or percentage of total requirements at a stated price during the period of the contract. In return, the buyer agrees to purchase at least a certain amount or fraction of total requirements from said seller.

These purchase contracts introduce a significant element of short-run stability into the prices actually paid by buyers and received by sellers. This element is reflected in prices collected from buyers; the WPI reports clearly tend to reflect not these prices, but the prices on

newly made contracts. Again a buyer's price gives a very different picture than the reported seller's price.

II

The widespread use of purchase contracts has implications for economic analysis extending beyond their immediate relevance to index number construction. Some of these implications are touched on here.

What functions do purchase contracts serve? An economist thinks immediately of hedges against price changes—analagous in function to future contracts in commodity markets. Some purchase contracts do serve a hedging function in this manner. Many do not. Many purchase contracts contain escalation clauses, which make the price vary in some specified relation to a specified trade journal published price. This is especially common with commodities in which seasonal variations are important, such as petroleum products. Other contracts allow price changes in accordance with the posted price of the seller—the change being mandatory in the event of a decrease and requiring a predetermined notification period in the event of an increase. The buyer usually has the option of canceling the contract in the event of a price increase.

Purchase contracts are not explicable simply as hedges against price increases, therefore. With that fact in mind, we may look at the advantages of a contract from the point of view of the buyer.

1. It provides protection against very short-run nonprice rationing by particular manufacturers or suppliers. Most purchasing agents remember situations in which particular products were simply not available on demand through the usual channels; e.g., steel in the immediate postwar period, copper in 1965 and 1966. The source of supply for marginal quantities then becomes the "grey market," often at very high prices.

2. It can economize on the costs of purchasing. These costs—excluding the invoice costs of the goods purchased—may be separated into two categories: transaction costs, defined as the costs attendant upon processing an order once a decision about the purchase has been reached; and information costs, defined as the costs of acquiring the information relevant to a decision of choice among suppliers. Transaction costs thus include such things as the costs of preparing orders, determining delivery dates, checking specifications, and the like. Information costs include the costs of acquiring information about suppliers and their products and about prices charged by particular suppliers.

Transaction costs, as defined here, are a function of the number of orders, not of the volume per order. Purchase contracts permit a re-

duction in number of orders without a corresponding increase in average inventory holdings, hence permit savings in transactions costs. Whenever a firm makes repetitive purchases of a standardized commodity, purchase contracts are feasible. Where the specifications vary from delivery to delivery, on the other hand, a blanket order requiring little further processing is simply not practical—because of the difficulty of specifying both quantities and prices for a large number of different items. A few figures, taken from preliminary data of the N.B.E.R. study, are illuminating in this regard. Chemicals, printing paper, petroleum products—all standardized products, with few variations in specifications by any one buyer—are bought on purchase contracts by over 50 percent of reporters to the N.B.E.R. using the items. The corresponding figures for metal mill products and electric motors, which require detailed descriptions, and are products for which orders are frequently switched from one type to another, are less than 15 percent in all cases.

The information relevant to purchasing operations is of two broad kinds. One concerns the physical characteristics of the products offered by different suppliers, suitability of those products for specific uses, and the reliability of the supplier with respect to delivery schedules, handling of rejects, and the like. These factors, pertaining to technical characteristics of the product, tend to change only slowly over time. The other relevant area of information—the one that economists are prone to think of first—is the price; and to a purchasing agent, the relevant price is the lowest one he can find, *ceteris paribus*, among the various prices quoted by suppliers. Information about relative prices of different suppliers as of one point in time may become obsolete fairly quickly—certainly more rapidly than information on technical characteristics.

Confronted with an array of suppliers with varying characteristics and prices, a rational purchasing agent acquires information on these facts from additional suppliers so long as additional information is worth more than the cost of acquiring it. Stigler[2] has shown that the expected returns from further search in terms of price are positive in that further search will yield a lower expected value of price and subject to diminishing returns in that the incremental drop in expected price decreases with additional search. Any search assures the buyer of a relatively favorable expected price—an advantage which can be preserved throughout the contract period at no further costs of search if a contract can be written at the low price thus found. (An escalation clause can allow for changes in the level of prices of all suppliers.) Presumably a buyer would increase the amount of search in the initial

[2] George J. Stigler, "The Economics of Information," *J.P.E.*, June, 1961, pp. 213–25.

period, for the volume to which the price then found is increased by use of the contract. Only if the relative prices of suppliers are not expected to change does the purchase contract not economize on information costs, for then the contract merely states what would exist in the absence of contracts.

To the extent that price depends on volume taken over some time rather than merely a single transaction, the purchase contract obviously allows a buyer to negotiate a lower price. That such economies may exist is attested to by both the statements of purchasing agents and of the recent cost theory of Armen Alchian[3]—each in its own argot. Purchasing agents report that purchase contracts are welcomed by sellers because they allow better production scheduling and lower selling expenses, while Alchian's Proposition I states that "the average and marginal cost per unit of total volume decreases as the total volume increases, holding the rate of production per unit of time constant."[4] As is often the case, the views of the theorist and of the practitioner differ in form of expression and agree in substance. This factor may be of importance to larger buyers, in their price negotiations concerning contracts.

The considerations cited suggest that for a particular commodity purchase contracts will tend to be used more by buyers of large quantities than by buyers of small quantities. Contracts can be useful as a way of economizing on costs of search, as a way of economizing on the

---

[3] Armen Alchian, "Costs and Outputs," in *The Allocation of Economic Resources* (Stanford Univ. Press, 1959).

[4] The variance among K series at a point in time i is

$$D = \frac{1}{K-1} \sum_{j=1}^{K} (X_{ij} - \overline{X}_{i\cdot})^2;$$

the mean of such variances over N periods is

$$\frac{1}{N(K-1)} \sum_{i} \sum_{j} (X_{ij} - \overline{X}_{i\cdot})^2.$$

This mean is algebraically equal to

$$\frac{1}{K(K-1)} \sum_{j<k} \sum \left\{ \frac{N-1}{N} \operatorname*{Var}_{i} (X_j - X_k) + (\overline{X}_{\cdot j} - \overline{X}_{\cdot k})^2 \right\}$$

$$= \frac{1}{2} \left\{ \frac{N-1}{N} \cdot (\text{mean variance of } (X_j - X_k)) + \text{mean of } (\overline{X}_{\cdot j} - \overline{X}_{\cdot k})^2 \right\},$$

where

$$\operatorname*{Var}_{i} (X_j - X_k)$$

is the variance of the difference between X_j and X_k taken over time,

$$\frac{1}{N} \sum_{i=1}^{N} [(X_{ij} - X_{ik}) - (\overline{X}_{\cdot j} - \overline{X}_{\cdot k})]^2,$$

with N in lieu of $N-1$ in the denominator. One component of the variance of D is thus the mean of all pairwise variances of differences—a useful measure of similarity of time series.

costs of transaction, and as a way of realizing certain economies in production. None of these is important in the case of a small or occasional buyer. Only if substantial volume is involved will the extra costs of the contract itself be justified. Again the preliminary data of the N.B.E.R. study are consistent with these assertions. When data on purchases by commodities are arranged in order of purchase volume, in all areas either none or a small proportion of the lowest quartile is covered by contracts, while the proportion in the top quartile is generally well over the proportion for the commodity as a whole.

III

In a number of commodities of the N.B.E.R. data there exists a noticeable variation in the timing of price changes among different reporters. This variability is apparent to the naked eye in series whose price tends to be stable for a number of months at a time, with shifts at various intervals. One sees roughly similar movements—a rise or fall of approximately equal magnitude—appearing in a number of series, but often displaced in time by one to six or more months from one another. This visual impression is borne out by a study of correlation coefficients among series from various firms on identical commodities. For each pair of series, correlation coefficients (used here purely as a descriptive measure) were computed with one series shifted forward six months, five months, four months, etc., and then backward one month, two months, etc. Marked effects on correlation coefficients were observed. Changes from .3 to .6, from .5 to .8, and similar changes could be observed by simply shifting price series in time relative to another by a few months.

It is clear, therefore, that a general movement of price does not affect all buyers at the same point or points in time. There are several possible reasons for this: (a) Selling firms do not change prices simultaneously, even when a corresponding change will be made by all. (b) Price concessions may be made by a seller to one or a few buyers at first, perhaps in response to lower offers made to these buyers by other sellers; only later are these concessions made available to all buyers. (c) Price changes will be delayed to some firms if purchase contracts without escalation provisions are in force. (d) In most industries other than steel, orders are billed at the price quoted at time of order, not at time of shipment; since the lag between order and delivery can and does vary, the data at which a change in price quotations are effective may differ among firms.

Each of these factors is known to have been involved in one or more timing differences noted in the N.B.E.R. data. All four factors cited would affect equally an index based on sellers' prices and an index

based on buyers' prices; only the first (lags in adjusting quoted prices) will affect an index based on the quoted prices or prices to new customers—such as the WPI.

These lags have important significance for the construction and use of index numbers. Two will be touched upon here.

1. The existence of lags of this nature points up certain difficulties in the comparison of time series—comparisons which are frequent in the construction of price indices, for purposes of describing the degree of similarity of price series going to make up a particular price index, and for purposes of comparing one or more price indices.

Because of these variable lags, it is not useful to compare first differences, or first differences of logarithms, to measure similarity. Two series can be highly correlated if one is displaced in time relative to another for all or part of the series, yet have very low correlation between first differences if the displacement is not made. Of course, one can always manipulate the series before using these measures, but the line between "adjusting" and "cooking" data becomes blurred here.

In some of the N.B.E.R. data, the correlation coefficient is very sensitive to such shifts in timing in actual data. This is another way of saying that if series which differ only in minor differences in timing of price changes are taken to be similar, then the correlation coefficient is not a useful statistic for describing similarity. An alternative measure —the variance of the difference between series—has the desirable property of being insensitive to such shifts in timing and can be generalized easily to consider a number of series simultaneously.[4]

2. Movements in the index constructed by taking an average (arithmetic or geometric) of the prices of several different series may display very different behavior in the short run than the behavior of any of the component series. Consider, as a simple example, a commodity whose price in each component series changes, on the average, once every six months. Suppose that these price changes do not occur simultaneously but differ by some months among reporters. Then the frequency of price change of the average will increase with the number of reporters, up to a maximum of one price change in the index for each reporting period—even though each individual series has but one price change every six months! As an arithmetic necessity, the mean price change in the index will be lower than the mean price change in the component series. It follows immediately that the use of the index for such purposes as studying price flexibility by such time-honored purposes as computing frequency of price change, mean amplitude of price change, and the like, is utterly fallacious with data such as these. The only valid way to compute such statistics is to compute them sep-

arately for each component series, and then combine the statistics themselves.

In a staff paper for the Price Statistics Review Committee, Harry McAllister[5] found that the number of price changes per period in the WPI, taken for many commodities, was an increasing function of the number of reporters. Since these prices are quoted sellers' prices, not received prices, it would appear that variability in time of sellers' quoted prices is substantial. This variability could in principle account for the observed results of both the N.B.E.R. data and the WPI data as analyzed by McAllister.

[5] Harry E. McAllister, "Statistical Factors Affecting the Stability of the Wholesale and Consumers' Price Indexes," Staff Paper No. 8 in N.B.E.R. Price Statistics Review Committee, *op. cit.*

ON THE MEASUREMENT OF PRICE AND QUALITY CHANGES IN SOME CONSUMER CAPITAL GOODS

By PHOEBUS J. DHRYMES

University of Pennsylvania

I. *Introduction*

This paper reports some preliminary and fragmentary findings of an investigation designed to determine whether it is feasible to correct for quality in the price of certain consumer capital goods by the methods suggested by Court [1] and more recently by Griliches *et al.* [2] [3].

If it is feasible to do so, then we can routinely construct price indices "corrected for quality change" and in this fashion provide a clearer indication of the true character of price movements. Price indices uncorrected for "quality" changes may seriously under- or overstate the extent of price movement. Of course, we should declare at the outset that by quality we shall mean the constellation of identifiable characteristics exhibited by a given product. In this study we are dealing with automobiles and refrigerators. Quality in the case of the former means, e.g., the weight, length, brake horsepower, etc., exhibited by a given model. In the case of the latter it means the height, weight, depth, freezer compartment capacity, etc., of the given refrigerator. Although it would be desirable, it is still not possible to include in the measurement of quality, the durability, or frequency of repair record, or the economy of operation of an automobile, or a refrigerator, due to lack of data.

Studies of this type are not new, as we have remarked above. What will particularly engage us is the examination of the question of homogeneity in the price behavior of various manufacturers. This is a problem of some consequence in the automobile industry, where well over 90 percent of the total car sales is accounted for by the output of three manufacturers. If the price decision-making process is appreciably different as between any two manufacturers, then it is clear that in attempting to "correct for quality" due attention must be paid to the division of the market by such manufacturers. Moreover, it would be inadmissible, under such circumstances, to estimate the pricing equation from a single cross-section, in which several manufacturers are represented. Rather we should estimate each manufacturer's decision rule separately. This is a problem that unfortunately has received little attention in the current literature.

II. *The Samples*

The automobile sample is in principle an exhaustive one for a certain subuniverse of models. The data were obtained[1] in the first instance from *Ward's* or *Automotive Industries*. In selecting models for inclusion the following criteria were employed: exclude sports cars, hardtops, convertibles and station wagons; include all two- and four-door models except deluxe versions of a given model not differing from the basic version in cylinders, weight, length, piston displacement, or horsepower. Notice that the criterion in the second criteria refers to standard equipment. Whenever data ambiguities arose the relevant manufacturer was contacted. The samples cover six years: 1953, 1957, 1961, 1962, 1963, and 1964. Two of the manufacturers provided corrections over the entire period, while one provided corrections for only one year.

The refrigerator sample was mainly obtained from *1966 Home Appliance Blue Book* and various issues of *Mart* magazine. By and large we used the entire set available from such publications. Some manufacturers were underrepresented or only incomplete information was available for their products. In such cases they were contacted directly and some information was solicited. Most responded. This sample covers the period 1950–65.

In the automobile sample we deal with 912 models (observations) of which G.M. accounts for 429, Chrysler for 303, and Ford for 180.

In the refrigerator sample we deal with 632 models (observations) of which Frigidaire accounts for 132, G.E. for 144, Hotpoint for 54, Kelvinator for 80, Philco and R.C.A. 38 each, and Westinghouse for 52. Other manufacturers represented are Admiral (9), Amana (4), Gibson (8), etc.

III. *Empirical Findings*

Automobiles. The work of Court [1], Griliches [2], and others has potentially important consequences for the construction of "quality corrected" price indices. The precise details of such application are subsidiary to the question whether it is in fact feasible to undertake quality correction routinely in this fashion. Two questions are of great importance, particularly in the case of automobiles. First, if we deal with a relation between list price and various identifiable physical characteristics of an automobile, then it would appear that such an equation is best interpreted as a cost plus desired markup relation.[2] We may thus

---

[1] The sample was collected in the first instance by I. Kravis and R. Lipsey. I am indebted to them for making it available to me.

[2] Other researchers have tended to interpret the parameters of such equations as expressions of consumer valuations of the physical characteristics of an automobile and thus as approximations of the implicit price attaching to such features. Of course such an interpretation should imply equality of coefficients in such equations for different manufacturers, provided the latter produce roughly similar products.

ask: is this costing function similar amongst different manufacturers? If not, it would appear that we ought to take into account the division of the market in constructing quality corrected price indices. Beyond that, if the answer to the previous question is no, then this entire procedure of quality correction could be open to serious doubt. Second, we might want to answer the following: is there some well-defined mathematical form that best describes this pricing function? Are the parameters of this function reasonably stable over time? Of course the parameters of such relations cannot be fixed over time since the cost structure of the relevant markets cannot be expected to remain fixed. Nonetheless we would not expect radical changes in the sign and magnitude of such coefficients from year to year.

We shall not be able to answer all these questions in this preliminary version, although some tentative answers can be suggested. Regarding the second problem, we have attempted to fit a linear, semilog or double-log form to the data, the results being reported in Tables 1, 2, 3, 4 (in the Appendix). The results reported there refer to cross-section regressions for the year 1961 for each manufacturer separately and for the entire sample. Since there is some question of confidentiality and the results are still tentative, we shall not identify the manufacturers.

In the tables W indicates weight in 10^2 lbs., L length in 10 inches, DIS displacement in cubic inches, BHP brake horsepower at 3800 revolutions per minute, MOD the number of units produced in the given year for the particular model and C, DOR, ATR, $P.S.$ are dummy variables assuming the value 1 if the model had as standard equipment eight cylinders, four doors, automatic transmission and power steering, respectively, and zero if not. Prices are stated in 10^2 dollars, and numbers in parentheses indicate t ratios. In addition to $\bar{R}^2$ we have given in the tables what we have called "antilog $\bar{R}^2$" in the case of semi- and double-log regressions. By that we mean the following. Let $\widetilde{\ln P}$ be the "predicted" logarithm of price from one of the last two equations of the tables, then obtain

(1) $$\tilde{P} = e^{\widetilde{\ln P}}.$$

Compute the sample correlation between $\tilde{P}$ of (1) and actual price. Thus "antilog $\bar{R}^2$" gives the correlation between actual and predicted price on the basis of the regression coefficients given in the last two columns of the tables. For purposes of comparison of the goodness of fit of the three functional forms this statistic imposes a reasonable degree of uniformity on the measures for all three forms and thus facilitates comparability. The first feature to note in these tables is that while the linear form is perceptibly inferior to the other two there is really not much to choose between the semi- and the double-log forms. Admittedly there is no simple (but rigorous) statistical test by which we could

support the·contention above, but in terms of operational expediency there is no doubt that we should reject the linear form.

The second feature worth pointing out is that the correlation coefficients are considerably higher when the relation is estimated on the basis of each manufacturer's sample than when the samples are pooled. This of course would lead us to believe that the pricing behavior of the various manufacturers differs appreciably. Fortunately, if we choose a specific functional form, the hypothesis above can be formally tested. For convenience let us choose the semilog form. It might appear that since displacement and $P.S.$ enter in the equation for manufacturer 1 and not in that of the others then *ipso facto* the coefficients must be different. However, this is an overhasty conclusion. First, we should note that for the sample of manufacturers 2 and 3 $P.S.$, when standard, was invariably accompanied by automatic transmission so that in fact the coefficient of ATR in Tables 2 and 3 measures the effect of both ATR and $P.S.$ on price. Second, for these same manufacturers BHP was highly correlated with displacement the coefficient for 1961 being about .96 and thus again the independent influence of these two characteristics on price could not be established. In fact, Griliches *et al.* [3] report the following relation, for 1961

(2) $$DIS = 73.4 + 1.07BHP, \qquad R^2 = .914$$

based on a cross-section of 36 models from various manufacturers. Since apparently the relation between displacement and horsepower is neither rigidly fixed nor freely variable, it would seem that some manufacturers in designing their models allow greater variability in the relation between those characteristics than others. For these reasons a formal test of the hypothesis of equality of coefficients should be formally carried out.

It can easily be proved that if we have s subsamples and if it is desired to test the hypothesis that the regression parameters characteristic of each sample are identical, we can proceed as follows.

Let

(3) $$Y_{\cdot i} = X_i \beta_{\cdot i} + u_{\cdot i} \quad i = 1, 2, \cdots, s$$

be the ith subsample; we assume $Y_{\cdot i}$ is $T_i \times 1$ so that each subsample may have a different number of observations; we also assume that

(3a) $u_{\cdot i} \sim N(0, \sigma_{ii}I) \; E(u_{\cdot i}u'_{\cdot j}) = 0 \quad i \neq j \quad i, j = 1, 2, \cdots, s.$

Let Q_i be the sum of squares of the residuals in the ith subsample— its parameters being estimated independently. Let Q_T be the sum of the squared residuals of the pooled sample. Then it can be shown that under the null hypothesis

(3b) $$\cdot_i = \beta, \qquad \sigma_{ii} = \sigma^2 \quad i = 1, 2, \cdots, s$$

(4) $$\frac{Q_T - \sum_{i=1}^{s} Q_i}{\sum_{i=1}^{s} Q_i} \cdot \frac{\sum_{i=1}^{s} T_i - sk}{(s-1)k} \sim F_{(s-1)k, \sum_{i=1}^{s} T_i}.$$

In our case $s=3$, $k=9$ and $\sum_{i=1}^{3} T_i = 192$; applying the above to our sample we conclude in the case of all three manufacturers that the F statistic is

(4a) $$F^* = 2.5594.$$

In view of the degrees of freedom involved (18, 167) the hypothesis is to be rejected even at the .5 percent level of significance.[3] It might be thought that manufacturers 2 and 3 may have identical parameters. This, however, is not borne out formally. A similar procedure applied to them alone yields the test statistic

(4b) $$F^* = 4.065.$$

Here the degrees of freedom are (8, 87) and thus the hypothesis is again rejected.

Thus it would appear that there are significant differences in the pricing behavior of the three manufacturers. This in conjunction with the fact that the coefficient of MOD turns out to be statistically significant would tend to support the interpretation that the equations estimated are indeed cost plus markup relations, and thus that the coefficients of the variables cannot reasonably be interpreted as approximations to the implicit price of such characteristics in the sense of an implicit consumer valuation. This being so, the question of whether one can use weights derived from such schemes to correct for quality variations becomes a serious one. It should be remarked that Griliches [2] attempted to do so, although his cross-sectional samples were perhaps not large enough to allow him to test the hypotheses we have tested above. Clearly additional work in this area will be needed to determine the general validity of our conclusions above. Such results will be incorporated in the final report of this study.

Let us illustrate the type of use we can make of our results above and the type of conclusions one is likely to arrive at given that the pricing behavior of various manufacturers differs appreciably.

Again, let us confine ourselves to the semilog form.

Let p_{ij} be the (log of the) price of a model with characteristics identi-

---

[3] Rejection is also implied for the double log and the linear form as well.

cal to those of the mean car in the sample for the jth year evaluated at the ith year's equation. In some sense p_{ij} is the cost (plus markup) of producing in the year i a model with the features of the mean model actually produced in the year j. It is clear that p_{jj} is then the (log of the) actual mean price of the sampled models in year j. Thus $p_{jj} - p_{ij}$ gives approximately that portion of the relative change in mean price between year i and year j which is not due to quality variations. We shall call this the "pure price effect." It is of course clear that $p_{jj} - p_{ii}$ gives the total (actual) relative change in mean price between the years i and j. It seems logical to define the quality effect or quality component of a price change by

$$(5) \qquad (p_{jj} - p_{ii}) - (p_{jj} - p_{ij}) = p_{ij} - p_{ii}.$$

Denote by

$$(5a) \qquad c_0 = p_{jj} - p_{ii}, \qquad c_1 = p_{ij} - p_{ii}, \qquad c_2 = p_{jj} - p_{ij}$$

so that c_0 indicates the actual price change, c_1 and c_2, respectively, the "quality" and "pure price" components.

The results of Table 5 may be given the following interpretation. The standard (average) model in the sample as a whole has experienced an apparent price decline between 1961 and 1964 of about 2.2 percent. But the intrinsic price behavior of automobiles is given by the "Pure Price" component under c_1. Thus over the same period this shows a decline of 7 percent; the discrepancy is made up by the fact that the quality of the average car has changed and this really should be thought of as equivalent to a 4.8 percent increase in price. The results are similar for manufacturers 1 and 2 but entirely different for manufacturer 3. Of course, we should remark that one of the variables, viz., *MOD*, cannot easily be thought of as a quality indicator, unless one is determined to so consider it; in this case we presumably could argue that the more units of a given model are produced, the less desirable is the car, much in the manner of women's dresses, where the value of the dress to the consumer is in part derived from the fact that there are not many (any?) others like it. In subsequent investigations, it would probably be best if that variable is removed from consideration.

We should further remark that the entries in Table 5 are not directly comparable to the automobile component of various published price indices; e.g., the consumer price index. In Table 5 all models receive equal weight; in the index, models are being weighted, as is sensible to do, by their sales volume or something equivalent. Finally, it should be remarked that the divergent conclusions emerging from consideration of manufacturer 3 or 2 or 1 tend to call into question the feasibility of routinely correcting for quality through the model pursued here and as suggested by others before. But I do not wish to be inflexible on this in view of the fragmentary and tentative character of the results.

Finally, before leaving this topic let us point out that the entries under c_2 in Table 5 may be tested for significance. It is clear that since the decomposition rests on the estimators for the parameters in each year the pure price effect will in truth be zero if the parameters have remained unchanged from year to year; an appropriate test statistic may be obtained and the test carried out.

Refrigerators. As in the case of automobiles we attempted to explain the price of refrigerators by the variation in certain identifiable physical characteristics. The characteristics considered were

CF = cubic footage capacity
FCF = freezer compartment cubic footage
H = height, inches
W = width, inches
D = depth, inches
$D1$ = dummy, 1 if meat drawer is available
$D3$ = dummy, 1 if egg shelf is available
$D4$ = dummy, 1 if butter shelf is available
$D7$ = dummy, 1 if shelves on freezer door are available
$D8$ = dummy, 1 if shelves are available in freezer compartment
$D9$ = dummy, 1 if ice ejector is available
$D13$ = dummy, 1 if shelves are sliding
$D14$ = dummy, 1 if shelves are swing-out
$D18$ = dummy, 1 if semiautomatic defrosting is available
$D19$ = dummy, 1 if automatic defrosting is available
$D20$ = dummy, 1 if automatic defrosting is available in the freezer
        section but not in the fresh foods section
$D21$ = dummy, 1 if completely frostless in freezer and fresh foods
        section
$D24$ = dummy, 1 if unit has 2 doors

Again, in this case we have attempted to fit linear, semilog, and double-log forms. We have found that while the linear form was appreciably inferior to the others, there was little basis for choosing between the semi- and double log. Here we shall base our discussion on the double-log form. The empirical results here are farther removed from finality than the automobile results. An analysis of annual cross-sections has not been completed. Furthermore, since no manufacturer had a sufficient number of models to permit identification of the parameters of the pricing equation of each individual manufacturer, we employed instead a mixed time series cross-section sample per manufacturer. The attempt here is to establish whether over the years the average pricing behavior of a given manufacturer differs from that of another.

In addition to that, we have obtained, on the assumption that pricing

behavior of manufacturers is homogeneous, estimates of the parameters of a pricing equation over the entire sample. If the hypothesis is correct, then this is indeed a relation reflecting the pricing behavior in this industry; if the hypothesis is not correct, however, the equation has no clear-cut interpretation, although one may be tempted to use it as an indication of average price behavior in the industry over the period in question.

In Table 6 we give the empirical results of this facet of the investigation. Perhaps a few comments will put this table in proper perspective. The first column refers to results obtained by considering the entire sample and is put there for completeness alone. This sample contains firms whose models are represented only at the early years of the sample (Crosley, Montgomery Ward) or only at later years (Gibson). The second column represents the results obtained when the sample considered contains only the manufacturers listed in the remainder of the table. Although the results do not differ markedly as to the magnitude of the coefficients in these two columns, nonetheless the goodness of fit is markedly inferior in the larger sample, and this is already an indication that the price policy of manufacturers will not be homogeneous. This is indeed borne out by examining the estimated parameters for Frigidaire, G.E., Hotpoint, Kelvinator, Philco, and Westinghouse. We note that the fit ($\bar{R}^2$) of these equations varies considerably from .6525 for Westinghouse to .8981 for G.E. There are also other indications of variability; thus for individual firms FCF appears to have an insignificant or negative coefficient while for the sample as a whole it appears to have a positive coefficient. Similarly H is unambiguously significant and has a positive coefficient in the total sample, while for the various manufacturer's subsamples it appears to be alternately significant with positive coefficient, insignificant or significant with negative coefficient. One can go on in this vein, but it does not particularly illuminate matters to take this view, since it is conceivable that multicollinearity problems may be particularly severe in one subsample and not in another or in the total sample. To this effect we revert to a formal technique and test the hypothesis that the parameters in all equations are identical by the method indicated above.

In this connection we should point out that the fact that zeros appear in the various places in the last six columns of the table does not necessarily indicate that the corresponding coefficients are (insignificantly different from) zero; rather it means that for the particular subsample the corresponding variable was identically zero and thus it was impossible to estimate such a parameter.

The test statistic implied by this sample is

(6) $F^* = 2.6885$

and in view of the degrees of freedom (77,406) the hypothesis is to be rejected unambiguously. Several pairwise tests also lead to rejection of similar hypotheses except in the case of Hotpoint and Westinghouse where the test statistic (1.47) lies within the acceptance region, but very close to its boundary.

We might well try to make some inference regarding the relative importance of quality variation in price changes over the years. However, since we do not have annual cross-sections, this is not quite satisfactory; so we confine our illustration to the total sample. The relevant information is given in Table 7. If the characteristics of the mean refrigerator in each year are evaluated according to the results of Table 6, there results the second column of Table 7; the latter thus gives the price variation over the years which is solely due to quality variation, since cost conditions are fixed. The quality component of price change is derived as the percent change obtained on the basis of this computed (cost constant) price and is given under q_1 in the table. Finally, the "pure price" component given under p_1 in the table is determined residually so that $q_1 + p_1$ equals the percent change in actual (mean) price. All such calculations have 1958 as a base year.

The conclusion emerging from the table is that apparent (actual) price behavior understates the extent to which the price of a quality corrected (average) refrigerator has fallen in recent years. While such calculations are useful as gross indications of the behavior of the relevant quantities, more definitive conclusions will have to be based on a more careful examination of annual cross-sections.

For in addition to other limitations indicated above, the results do not adequately allow for the fact that the general price level has been fluctuating over the sample period. To some extent this can be allowed for in this aggregative context by introducing time as a variable. It is not clear that this is a very satisfactory substitute for annual cross-sections. Due to this as well as time and space limitations, this will not be pursued here.

IV. *Conclusion*

We have attempted in this paper to give a fragmentary account of certain results pertaining to the feasibility of constructing quality corrected price indices on the basis of suggestions that regression of price on various identifiable characteristics of the relevant products adequately represents the effect of quality in price.

Our tentative conclusion is that it is very difficult to apply such techniques routinely. First, there is the problem that the pricing equations for various manufacturers are not homogeneous, so that the same item is differently evaluated in the different pricing equations and thus

it is not clear what is its quality contribution to the price of the product. Second, there is some evidence that the parameters of such equations vary significantly from year to year. What, for instance, would be our interpretation if, for a given manufacturer of automobiles, length is significant with a positive coefficient for one year but for the next—and still the same manufacturer—length is insignificant, or significant but with a negative coefficient?

Particularly in the case of automobiles, fads carefully nurtured by the industry's advertising are common, and since we do not allow for them in our estimation procedure, we are likely to be led astray if we are totally objective about our quality correction.

REFERENCES

1. A. T. Court, "Hedonic Price Indexes with Automobile Examples" in *The Dynamics of Automobile Demand* (New York, 1939).
2. Z. Griliches, "Hedonic Price Indexes for Automobiles; . . . " in *The Price Statistics of the Federal Government* (N.B.E.R., 1961).
3. F. M. Fisher, Z. Griliches, and C. Kaysen, "The Cost of Automobile Model Changes since 1949," *J.P.E.*, 1962, p. 433.

<div align="center">

APPENDIX

TABLE 1

ALTERNATIVE FUNCTIONAL FORMS, Mf 1, 1961

</div>

	Linear $P = \sum_{i=1}^{n} \alpha_i X_i$	Semilog $\ln P = \sum_{i=1}^{n} \beta_i X_i$	Double Log $\ln P = \sum_{i=1}^{n} \gamma_i \ln X_i$
W	.7898 (1.2450)	.0171 (1.5375)	−0.3777 (−0.9555)
L	4.7060 (1.7042)	.0798 (1.6456)	4.1890 (4.2241)
DIS	−0.1114 (−4.8789)	−0.0017 (−4.2320)	−0.5311 (−3.7497)
BHP	0.0477 (2.7863)	0.0012 (3.9164)	.3339 (4.6638)
MOD	−0.0579 (−3.8918)	−0.0013 (−4.9417)	−0.1152 (−6.4716)
C	1.5376 (.6602)	.0337 (.8236)	.0622 (.0492)
DOR	−2.6438 (−1.6201)	−0.0516 (−1.8018)	−0.0678 (−2.3798)
ATR	−1.4994 (−0.7134)	−0.0058 (−0.1562)	.0193 (.5216)
$P.S.$	12.0165 (3.6087)	.3162 (5.4091)	.3085 (5.3208)
Const.	−71.3150 (−1.9571)	1.3050 (2.0402)	−6.4053 (−3.4792)
$\bar{R}^2$	.8226	.9069	.9093
Antilog $\bar{R}^2$	.8226	.8980	.9135

TABLE 2

ALTERNATIVE FUNCTIONAL FORMS, Mf 2, 1961

	Linear $P = \sum_{i=1}^{n} \alpha_i X_i$	Semilog $\ln P = \sum_{i=1}^{n} \beta_i X_i$	Double Log $\ln P = \sum_{i=1}^{n} \gamma_i \ln X_i$
W	.7412 (2.0924)	.0311 (2.1342)	.9882 (2.1421)
L	−1.2551 (−1.1584)	−0.0440 (−0.9881)	−1.1496 (−1.2486)
BHP	.0192 (3.3766)	.0008 (3.3181)	.1609 (2.8456)
MOD	−0.0893 (−3.1684)	−0.0033 (−2.8530)	−0.0915 (−3.5895)
C	−1.7973 (−2.2625)	−0.0693 (−2.2186)	−0.0666 (−1.9375)
DOR	−0.5699 (−1.3910)	−0.0282 (−1.6732)	−0.0298 (−1.7254)
ATR	7.5232 (5.3169)	0.2005 (3.4493)	.2130 (3.5460)
Const.	23.3652 (2.0099)	2.9577 (6.1929)	2.6271 (1.9202)
$\bar{R}^2$	.8645	.8645	.8557
Antilog $\bar{R}^2$	.8645	.8702	.8635

TABLE 3

ALTERNATIVE FUNCTIONAL FORMS, Mf 3, 1961

	Linear $P = \sum_{i=1}^{n} \alpha_i X_i$	Semilog $\ln P = \sum_{i=1}^{n} \beta_i X_i$	Double Log $\ln P = \sum_{i=1}^{n} \gamma_i \ln X_i$
N	.7176 (3.1004)	.0326 (3.2060)	.9532 (3.1372)
L	−2.0737 (1.8343)	−0.0868 (−1.7464)	−1.8124 (−1.8424)
BHP	.0027 (.5913)	.0001 (.4771)	.0341 (.7301)
MOD	−0.0189 (−2.8428)	−0.0008 (−2.7072)	−0.0573 (−2.8506)
C	.4287 (.7568)	.0192 (.7703)	.0204 (.7750)
DOR	−0.1516 (−0.3560)	−0.0075 (−0.4009)	−0.0106 (−0.5790)
ATR	21.5954 (6.6018)	.3917 (2.7242)	.2064 (1.0054)
Const.	41.5015 (2.5160)	3.8156 (5.2620)	5.2898 (2.5447)
$\bar{R}^2$	.9731	.9468	.9465
Antilog $\bar{R}^2$	.9731	.9732	.9732

TABLE 4

ALTERNATIVE FUNCTIONAL FORMS, TOTAL 1961

	Linear $P = \sum\limits_{i=1}^{n} \alpha_i X_i$	Semilog $\ln P = \sum\limits_{i=1}^{n} \beta_i X_i$	Double Log $\ln P = \sum\limits_{i=1}^{n} \gamma_i \ln X_i$
W	0.8861 (3.5491)	0.0247 (5.2114)	0.3763 (2.2063)
L	−1.0440 (−1.0123)	−0.0300 (−1.5305)	0.3071 (0.6918)
BHP	−0.0056 (−0.8206)	0.0002 (1.6677)	0.0614 (1.8327)
MOD	−0.0115 (−1.9044)	−0.0003 (−3.0399)	−0.0204 (−2.6775)
C	0.7605 (0.7484)	0.0385 (1.9944)	0.0462 (2.03552)
DOR	−1.7007 (−2.0621)	−0.0430 (−2.7458)	−0.0513 (−3.0666)
ATR	0.2325 (0.1504)	0.0363 (1.2366)	0.0720 (2.3577)
$P.S.$	20.8004 (9.9473)	0.4453 (11.2165)	0.5067 (12.3957)
Const.	16.3521 (1.9747)	2.8690 (10.8544)	0.6121 (0.7405)
$\bar{R}^2$	.7762	.8823	.8655
Antilog $\bar{R}^2$	.7762	.827	.790

TABLE 5

"QUALITY" AND "PURE PRICE" COMPONENTS OF CUMULATIVE PRICE CHANGES (PERCENTAGES)—U.S. 1953-64

	Total			Mf 1			Mf 2			Mf 3		
	c_0	c_1	c_2	c_0	c_1	c_2	c_0	c_1	c_2	c_0	c_1	c_2
1953	−4.3	4.25	−0.05	−13.1	−10.6	− 2.5	4.2	−5.7	9.9	n.a.	n.a.	n.a.
1957	5.2	−2.8	8.0	− 0.6	− 2.9	− 2.3	8.9	2.5	6.4	n.a.	n.a.	n.a.
1961	0	0	0	0	0	0	0	0	0	0	0	0
1962	−0.6	5.0	−5.6	2.2	7.9	− 5.7	−1.2	4.3	−5.5	−1.7	−3.3	1.6
1963	.3	2.5	−2.2	3.4	7.6	− 4.2	−0.1	6.0	−6.1	−3.3	−5.1	1.8
1964	−2.2	4.8	−7.0	− 1.6	9.2	−10.8	1.7	10.0	−8.3	−2.6	−6.6	4.0

TABLE 6
PRICING EQUATION
MAJOR MANUFACTURERS
(Double Log) U. S. 1950–65

	622 obs.	500 obs.	Frigidaire 132 obs.	G.E. 144 obs.	Hotpoint 54 obs.	Kelvinator 80 obs.	Philco 38 obs.	Westinghouse 52 obs.
FCF	.0597 (2.653)	.0461 (1.795)	−0.0276 (.412)	.0165 (.441)	.0897 (.994)	.0948 (1.324)	.2178 (1.313)	−0.2724 (−1.7333)
H	.1801 (2.442)	.2460 (3.199)	.5473 (2.565)	.0855 (.626)	.5726 (2.261)	.3853 (2.960)	−1.6334 (−2.652)	.4353 (.543)
W	.5475 (8.117)	.5757 (7.933)	.5138 (2.529)	1.1234 (7.294)	.2700 (.974)	.5046 (4.972)	1.9455 (2.501)	.7548 (1.298)
D	−0.4958 (−5.186)	−0.5981 (−5.511)	−0.9238 (−3.717)	−0.934 (−5.011)	−0.6587 (−1.284)	−0.6853 (−3.570)	.3389 (.633)	−1.0064 (−1.358)
D1	.1164 (6.759)	.1020 (5.473)	.0741 (2.112)	.0413 (1.417)	.1891 (2.817)	.1101 (2.698)	.0353 (.367)	.0700 (.715)
D3	−0.0690 (−3.241)	−0.0834 (−3.333)	−0.0750 (−1.164)	−0.0950 (−2.792)	−0.0145 (−0.138)	−0.1147 (−1.815)	−0.1287 (−1.517)	−0.1266 (−0.754)
D4	.0594 (2.546)	.0645 (2.552)	.1029 (2.122)	.0605 (1.578)	−0.6139 (−0.148)	.0133 (.236)	.1071 (1.153)	.1994 (1.461)
D7	−0.1817 (−4.938)	−0.2109 (−5.507)	−0.0953 (−0.623)	−0.2301 (−3.935)	−0.3741 (−3.622)	−0.1738 (−1.959)	−0.1994 (−1.685)	−0.2565 (−1.350)
D8	.1783 (7.3555)	0.1925 (7.2111)	.1149 (2.473)	.2237 (5.572)	.0301 (.334)	.1301 (1.874)	.2578 (1.958)	.2947 (2.133)

TABLE 6—Continued

D9	.12309 (3.780)	.1241 (3.824)	.0228 (.536)	-0.1895 (-1.414)	0	0	.2631 (1.270)	.1389 (1.007)
D13	.0414 (2.074)	.0401 (1.817)	.2015 (4.803)	-0.0133 (-0.269)	.1031 (1.464)	.0113 (.276)	-0.0113 (-0.111)	.1420 (1.242)
D14	.2032 (6.272)	.2128 (6.336)	0	.1996 (4.048)	.3976 (3.251)	0	0	0
D18	.1407 (4.315)	.1058 (2.998)	.1010 (1.897)	-0.0337 (-0.433)	0	.1738 (3.002)	0	.4340 (2.504)
D19	.3333 (11.275)	.3225 (9.334)	.2901 (4.281)	.2420 (4.839)	.1472 (1.191)	.2909 (3.952)	.3151 (2.302)	.3385 (2.113)
D20	.2853 (9.158)	.2884 (8.905)	.2501 (5.397)	.1508 (2.493)	.4240 (3.144)	.5489 (17.175)	.1506 (1.165)	.4757 (2.232)
D21	.2804 (6.957)	.3189 (7.635)	.3566 (5.817)	.2896 (4.194)	.4513 (2.703)	.6209 (6.238)	.1624 (.817)	.2920 (1.278)
D22	.1406 (3.536)	.1394 (3.336)	.0867 (.576)	.2482 (3.754)	-0.1479 (-1.177)	-0.1291 (-1.228)	.1929 (1.504)	.2746 (1.030)
R^2	.7756	.8011	.8668	.8981	.6842	.8755	.8032	.6525

TABLE 7

QUALITY AND PRICE BEHAVIOR
REFRIGERATORS, ALL MANUFACTURERS
U.S. 1950–65

	Actual Price	Computed Price	Actual Price Change Relative to 1958 %	q_1	p_1
1950	284.	293.	−34.0	−24.0	−10.0
1951	318.	277.	−22.0	−29.5	7.3
1952	371.	369.	− 6.9	− 8.5	1.6
1953	365.	321.	− 9.9	−13.9	4.0
1954	340.	305.	−16.5	−20.0	3.6
1955	368.	303.	−11.0	−15.1	4.1
1956	375.	341.	− 9.9	− 8.8	− 1.1
1957	407.	356.	− 2.3	−15.6	16.6
1958	418.	371.	0	0	0
1959	377.	338.	−10.5	− 9.6	− 0.9
1960	483.	418.	16.9	11.5	5.4
1961	392.	369.	− 7.2	− 9.5	2.3
1962	356.	329.	−16.2	−11.8	− 4.4
1963	411.	387.	− 1.1	3.1	− 4.2
1964	422.	391.	1.1	5.1	− 4.0
1965	364.	416.	−12.0	11.9	−23.1

DISCUSSION

GREGORY M. KIPNIS: The presented papers clearly represent an attempt to push a little further out both the positive theoretical and empirical frontiers of the field of price measurement. Messrs. Kravis, Lipsey, and Kindahl are significantly adding to the pool of potentially usable new price data by demonstrating the feasibility of collecting and/or using such data. On the other hand, Dhrymes has given us reason to doubt that hedonic price indexes can be "routinely" constructed.

In passing, I would like to note that the name of this session might be appropriately renamed from "Measurement of Price Change" to "Measurement of Price Change—However Defined." This quibble, I believe, would better serve to draw a distinction between the statistical problems of manipulation and measurement and the conceptual problems faced in defining price.

It follows, from this latter point, that because of the differing uses and concepts for which price data are employed, we do not all have the same thing in mind when we speak of price. Thus, it becomes important that we specify precisely what we mean by price in each use!

Let me illustrate. We use price data to measure inflation, cost of living, price flexibility, and price competitiveness. At the same time, price indexes are used as deflators for the national income and product accounts (which in turn affect measures of productivity).

It is clear that in each case we are talking about a different kind of price measure. The differences arise, in part, from technical and methodological differences (e.g., extent of coverage and base year versus current year weights), differences in the market level at which data are collected (e.g., retail versus wholesale) and the extent to which adjustment is made for nonmarket-priced characteristics (e.g., quality and productivity differentials).

Regarding the last point, the problem of adjustment for changes in quality and productivity pose some of the thorniest conceptual problems. For example, in deflating the components of gross national product there are a number of price and pseudo-price adjustments which can be made depending on the desired result.

First, in the case of current dollar expenditures for nondurables, adjustment may be made for changes in quality over time to reflect the nonmarket-priced services received. In so doing, one estimates the real value of services vis-à-vis real cost. In the case of durable expenditures, adjustment may be made for productivity changes to reflect differences in the capacity to produce.

Next, expenditures for durables and nondurables may be deflated, removing the influence of changes in market price, without regard to changes in quality or productivity. This procedure provides a measure of the constant dollar cost of output.

Last, adjustment for both changes in quality, productivity and market price may be made to reflect the flow of real constant dollar services received.

The approach deemed most desirable, from the point of view of the national accounts, was thoroughly and extensively argued by Edward Denison.[1] Specifically, expenditures for producers' durable equipment should be deflated with price data unadjusted for changes in productivity in order not to bias subsequent estimates of capital productivity and the capital output ratio. In the case of consumer durable expenditures, however, price adjustment, where meaningful, should be made which includes quality and/or productivity changes. The logic underlying this distinction is that in the case of the former we are dealing with intermediate products which become the instruments of production, whereas the latter are final goods.[2]

Changing our focus from the price adjustment process as it affects estimates of real income and product, consider that an alternative route would be to adjust the price indexes themselves for quality and productivity differences. The resulting price index, for lack of a better term, could be referred to as a measure of real price movement.

These rather briefly stated illustrations should have made clear that there is no unique measure of price. In fact it is the use which defines the measure to be used. Further, whatever operations are performed, the resulting price index, while conceptually meaningful, is clearly restricted in the alternative uses to which it can be put.

In this connection, any attempt at understanding the nature of price behavior of hybrid indexes such as the Implicit Price Index (i.e., the implicit price deflator for GNP) which combines price data which differ both in concept and market level is rendered meaningless. The data include retail, wholesale, agricultural and producer prices, earnings and implicit price data, and both productivity and quality adjusted and unadjusted data.[3]

Returning to the papers presented, we find Kindahl concerned simultaneously with the problem of measuring the timing of price change impact and the actual prices at which industrial goods are transacted in volume. Data on buyers' prices, being theoretically consistent with his goals, serve his purpose well. However, the suggested use of sellers' prices for purposes of deflation of value of output would be less than ideal. Sellers' prices will reflect changes in both the sales price, net of discounts and the freight

---

[1] Edward Denison, "Theoretical Aspects of Quality Change, Capital Consumption, and Net Capital Formation," *Problems of Capital Formation*, Vol. 19, Studies in Income and Wealth (N.B.E.R., 1957), pp. 215–60.

[2] To the extent that expenditures for consumer durables (e.g., automobiles, appliances, and laundry equipment) are counted as part of net capital formation and the attendant estimates of imputed income are made, then price adjustment should be free of productivity and quality changes. The reality of this consideration is made quite clear in the yet unpublished, "Household Capital Formation: Growth, Cyclical Behavior, and Financing, 1897–1962," by F. Thomas Juster.

[3] Unique data and an extensive description of the techniques of deflation, mathematical properties of the implicit price index, and the weights associated with each of the classes of price data embodied in the implicit price deflators can be found in my paper published as Appendix C in *Inflation and the Price Indexes*, Studies in Business Economics, No. 94 (Nat. Ind. Conf. Bd., 1966), by Martin R. Gainsbrugh and Jules Backman.

equalization costs picked up by the seller. Resultant productivity measures would be biased in the opposite direction of relative changes in transportation costs; i.e., if the relative change in transportation costs is greater than the relative change in the cost of the industrial commodity, productivity estimates would be biased downward. If, on the other hand, buyers' prices were used, the bias would run in the opposite direction.

It is expected that the industrial price study will shed new light on the magnitude and importance of this component of price change.

With great elegance Dhrymes has succeeded in demonstrating that quality change in consumer durables cannot be measured by regressing identifiable physical characteristics against list price, if manufacturers employ a cost-plus markup pricing equation. The parameter estimates, if the equation were correctly specified, would indicate the relative effect on total cost of each of the considered physical characteristics. However, if the pricing equation for each characteristic were based upon estimates of demand elasticity, then the parameter estimates could be used to adjust price for quality differences as viewed by the consumer. Since the nature of the price equation employed likely differs from firm to firm, this would clearly argue against routine construction of hedonic price indexes.

INVITED DOCTORAL DISSERTATIONS I

DAILY NEWSPAPERS, MONOPOLISTIC COMPETITION, AND ECONOMIES OF SCALE*

By JAMES N. ROSSE

Stanford University

The study reported in this paper found cross-section econometric evidence of substantial economies of scale among smaller daily newspaper firms, found additional evidence that such economies extend to all scales of newspaper production, and found some evidence that such economies have not changed over time. These findings provide a necessary basis for understanding the isolated nature of the daily newspaper industry's structure.[1] Novelties of the study lie in its econometric use of a Chamberlinian model and in its indirect methods of estimating cost function parameters.

An Industry of Isolated Competitors

In 1963, 96.6 percent of the 1,493 U.S. cities having at least one daily newspaper contained only one producing establishment. But Markham has found evidence of substantial intermedia (and some intramedia) competition [12]. Application of monopoly theory to the newspaper firm would therefore depend on an unreasonably narrow definition of relevant product market.

Chamberlin's theory of monopolistic competition [1] is a more appropriate vehicle for investigation of the industry; indeed, the newspaper is a prime example of Chamberlin's variable product [2]. [3] [5] [12] [15] and [16] have made explicit or implicit use of Chamberlinian theory in study of newspaper firms.

Within a Chamberlinian context, the term "isolated" characterizes the industry's structure better than "local monopoly." A typical member of the industry is isolated in the sense that its demand cross-elasticities with respect to prices charged by competing media are certainly finite. On the other hand, it is distinguished from a monopoly

[1] The word "industry" is loosely used to obviate need for referring to "that set of firms which sell products commonly called daily newspapers."

by the fact that not all demand cross-elasticities are negligibly close to zero.[2]

The full force of Kaldor's "nearest competitor" criticism [9] of Chamberlinian theory does not fall on a theory of isolated monopolistic competitors. Because of the absence of any very "near" competitors in the Kaldorian sense and because of the multiplicity of not so "near" competitors (local weekly and remote daily newspapers, electronic and other print media), an isolated newspaper firm may be regarded as viewing demands for its products as parametrically given and the core of Chamberlinian theory retained.

Why Are Newspaper Firms Isolated?

New York City contained twelve newspaper firms in 1963. If multiple firms can exist there, why not in smaller cities as well? In 1923 and 1963, respectively, 38.7 percent and 3.4 percent of U.S. cities with daily newspapers contained more than one. If multiple firms could exist in 1923, why not today?

The existence of economies of large-scale production is a necessary condition for answering these questions within the context of competitive economic theory [2, p. 11]. If not all consumers of newspaper products have identical product preferences and if per unit costs of small firms are no higher than those of large firms, then there should be at least one newspaper for every identifiable set of buyer tastes or needs; each buyer could obtain his most preferred newspaper product at no difference in price.

If scale economies exist, however, newspaper buyers in the aggregate must choose between multiple "tailor-made" newspaper products produced by small relatively high-cost firms and a cheaper "common-denominator" product produced by a large-scale firm. These choices or product preferences will, of course, be reflected in demands facing each firm. For sufficiently strong buyer product preferences and/or for sufficiently small differences in per unit cost (as in large markets) multiple firms may survive; otherwise, scale economies may well overbalance diversity in product preference to create an isolated firm. The less isolated industry structure of 1923 may have become 1963's isolated structure if preferences became more homogeneous or if scale economies became more pronounced.[3]

---

[2] Any attempt at precise definition of the descriptive term "isolated" breaks down on the question of "how large is negligibly small?" McFarland and Olson have attempted to resurrect monopoly theory but run into difficulty on this question [14].

[3] A general downward shift of demand for newspaper products or a general upward shift in costs could also cause the industry's structure to become more isolated. Evidence not reported here suggests that a weakening of buyer preference for tailor-made newspaper products caused by changing advertiser needs, by increased availability of alternative communication media, and by increased homogeneity of the reading public has been the most important influence on the industry's structure, but the evidence is not conclusive.

Quantifying Newspaper Output

Following suggestions implicit in [2] [3] and [9], I have measured newspaper output as a vector quantity. The output of a firm publishing once each weekday has components measuring circulation, advertising space, nonadvertising or news space, product delivery, and product quality. It seems reasonable to measure the first three by paid circulation and column inches of advertising and nonadvertising space per issue.

Representation of the delivery component is needed before outputs can be compared since the newspaper is a delivered product. Circulation and market area seem the most important determinants of delivery cost. Circulation is already measured, so subscriber market area completes quantification of the delivery component.

Newspaper quality is given a narrow interpretation as in [3]. It is interpreted not by abstract standards but rather relative to the firm's perception of consumer preferences and is measured by the cost of the factor combination needed to obtain it. If firms are price takers in markets for editorial inputs, editorial cost per column inch of news space may be a useful although imperfect index of product quality.

Scale Economies Defined

For a scalar output process, one may define scale economies as the existence of declining average cost.[4] The joint-variable-product characteristics of newspaper output make definition more complicated.

If cost per subscription declines as circulation increases, *ceteris paribus*, scale economies in circulation are said to exist. Similarly, if cost per advertising (or news) inch declines as column inches of advertising (or news) space increase, *ceteris paribus*, scale economies in advertising (or news) space are said to exist.

Examples can be constructed showing that, for some configuration of buyer preferences, scale economies by any one of these definitions might explain the industry's isolated structure. Scale economies without additional specification may refer to one or all of the above. Unless otherwise stated, scale economies are defined relative to long-run cost.

A variable or differentiable product, such as the daily newspaper, produced under a regime of scale economies creates a variant in the public good problem of welfare economics (Samuelson [17]); not only does the quantity available to one consumer at a given price depend on demands of all other consumers, but the kind or quality of product he can obtain depends on their product preferences.

[4] More fundamental definitions of scale economies can be constructed (e.g., Debreu [4, para. 3.3]). Because of the empirical orientation of this study and because of the difficulty in specifying empirically useful joint-variable-output production functions (however, see Mundlak [13] and Klein [10]), I have chosen to stay entirely within the context of cost functions. Sufficient conditions for the existence of needed cost functions can be found from Uzawa [18] by extending his duality proofs to take account of vector valued outputs. To be consistent with a realistic model of the newspaper firm, Uzawa's input supply functions would need to be modified as well.

If scale economies exist, the rents earned in long-run equilibrium by such firm-specific inputs as location would be capitalized and might appear on a firm's books as cost [7] [19, p. 41]. If such costs were uniform for all firms, they would present no problem in definition and observation of newspaper cost, but the existence of scale economies and varying city size imply the absence of such uniformity for isolated firms. I therefore define newspaper cost net of such firm-specific rents that may exist.

An Indirect Method of Estimating Cost Function Parameters

Problems of identifying and observing "real" cost are handled by an indirect method inspired by Friedman [7]. Parameters of newspaper cost functions are inferred from cross-section observed outputs and output prices under an assumption of profit maximizing behavior on the part of individual firms. Such an approach runs into difficulty in a perfectly competitive industry, but, as Walters notes [19, p. 41], it may be fruitful in a monopolistically competitive industry where firms face differing although parametrically given demand functions.

The method requires detailed specification of demand and cost functions common to all firms. To keep model building as simple as possible and still consistent with realistic assumptions, I have restricted most empirical studies to small isolated firms producing a single weekday edition. About 80 percent of firms in the industry fall within the output range considered.

The First Empirical Study

Data for an empirical test of scale economy hypotheses were generated from published sources and from direct inquiry. The sample included 171 observations from 59 firms at two-year intervals from 1958 to 1964. The firms were nonmetropolitan middle western isolated firms publishing a single weekday edition of four to fifty thousand copies.

A stochastic short-run (annual) model for firms in the sample was constructed by specifying a demand function for each of circulation and advertising and a marginal cost function for each of circulation, advertising space, and news space. The demand functions imply marginal revenue functions not only for circulation and advertising but also for news space since demand for circulation depends on the amount of news space published.

Profit maximizing choices by each firm of its short-run instrument variables (advertising and circulation prices and quantity of news space) must satisfy three marginal revenue–marginal cost equalities in addition to the two demand equations. Thus five endogenous variables (news space quantity and both prices and quantities of circulation and advertising) are determined by five stochastic equations; parameter esti-

mates of marginal cost functions may be inferred from parameter esti-
mates of the complete system.

Product quality, area of the firm's subscription market, scale of plant
(as indexed by undepreciated book value of non-real-estate physical
assets), and color printing capability would be endogenous variables in
a long-run model but were taken as fixed by long-run decision and there-
fore exogenous in the short-run model. Four other exogenous variables
were the number of households and the number of competitive adver-
tising media in each firm's subscription market and income and retail
sales per household.

Parameters of demand and marginal cost functions were assumed
identical for all firms and for all four years except that separate constant
terms were calculated for each year. Since estimation of marginal cost
parameters depends critically on evaluation of derivatives of estimated
demand function parameters and since substantial collinearity was
present in the observation matrix, results are quite sensitive to errors in
specification of algebraic form. This kind of error was minimized by
subjecting repeated trial specifications of the model to test against
formalized and theoretically derived a priori acceptance criteria.

Demand functions were linear in logarithmic values and marginal
cost functions were linear in arithmetic values in the reported model.
With this specification, the model included nonlinear identities and full-
information estimators could not be calculated. A three-stage procedure
was used where limited-information estimators of demand function
parameters were calculated first. These were then used to calculate
implied marginal revenues for each observation, and the results of this
calculation were used in the three equations characterizing necessary
conditions for profit maximization from which limited-information esti-
mators of marginal cost parameters could be obtained.

Sufficient a priori conditions for identifiability of the model were
established using Fisher's criteria [6] but *ex post* identification tests
were ambiguous. Reported results are based on limited-information
maximum likelihood estimators for four equations and two-stage least
squares estimators for the fifth. Hypothesis test statistics were calcu-
lated using the known asymptotic distributions of these estimators.[5]

Results of the First Empirical Study

Hypotheses were tested by examining estimates of the second dif-
ferential of total cost. By the definitions of scale economies, all variables
except plant scale and outputs of advertising, news space, and circula-
tion must be held constant when testing hypotheses. Since assumed

[5] Econometric detail is reported more completely in a forthcoming article titled, "On
Estimating Cost Function Parameters Without Using Cost Data: Illustrated Methodology."

marginal cost functions are linear, the quadratic form defining the second differential of total cost will have a constant matrix. Holding product quality, market area, color printing capability and time variables constant is equivalent to neglecting rows and columns of the matrix corresponding to these variables. The remaining 4×4 submatrix contains the sixteen constant second partial derivatives of cost with respect to the four variables of interest. Estimates of fifteen of these constants can be found from parameters of marginal cost and the direct

TABLE 1

An Increment of a Normalized* Unit of	Gives Rise to an Estimated Second Differential of per Issue Cost of	and the Hypothesis	May be Rejected with a Probability of Error of Approximately
1. Advertising output per issue..	−286	$0 \leq d^2C$	.066
2. News space output per issue..	−321	$0 \leq d^2C$	.087
3. Both advertising and news space per issue............	−459	$0 \leq d^2C$	.031
4. Circulation output per issue..	−356	$0 \leq d^2C$	.174
5. Both advertising and plant scale...................	−289	$0 \leq d^2C$	.052
6. Both news space and plant scale...................	−429	$0 \leq d^2C$	.031
7. Advertising, news space, and plant scale..............	−570	$0 \leq d^2C$	.00001 −
8. Both circulation and plant scale...................	+168	$d^2C \leq 0$	.367

* Normalized units correspond to average sample values. One normalized unit is equivalent to 1,640 column inches of advertising, 1,656 column inches of news space, per issue circulation of 15,146, or a plant scale asset value of $456,570.

second partial with respect to plant scale was assumed zero completing the matrix.[6]

When the second differential of total cost is negative for specified output increments, the short-run marginal cost function defined by those increments is negatively sloped. Provided costs fixed relative to those increments are nonnegative, the corresponding average cost function is negatively sloped as well, leading to an inference of short-run scale economies.

It seems evident from items 1–3 in Table 1 that short-run scale economies exist in production of advertising and/or news space. Item 4 suggests the same result for circulation, but the conclusion must be qualified by its large error probability and by knowledge that some unmeasured deterioration in product quality may be taking place simultaneously.[7]

[6] Since plant scale was measured by the dollar value of assets, this assumption is not unreasonable.

[7] Larger circulation with no increase in press capacity means longer press runs and less timely news in the final product.

Again choosing arbitrary output increments, if a given plant s minimizes cost and if an increment of plant scale exists for which second differential of total cost is negative, the long-run marginal function defined by the output increments is negatively sloped. Its responding long-run average cost function displays scale economies

It seems evident from items 5–7 of Table 1 that long-run scale ec mies exist in production of advertising and/or news space. The estim. second differentials were negative for all increments of plant scale are significantly less than zero for increments proportional to "norr values.

When a circulation increment is considered, the estimated se(differential remains negative for small increments of plant scale as item 8 shows, the estimate is positive for proportional plant s increases. The probability of error in rejecting a nonpositive differe1 hypothesis is high, however.

Considering items 4 and 8 and considering what is known of n paper technology, I have concluded that a hypothesis of nearly cons marginal cost in circulation is not unjustified. The slight decline in sl run cost can be attributed to longer press runs at little or no increa: pressmen's work hours, but the unitized nature of press equipr suggests that increments of press capacity bring proportional cost creases. If this is correct, average circulation cost should be decli: since there are substantial costs fixed relative to the circulation out

A Second Empirical Study

The above conclusions are limited by the relatively small size firms in the sample used. Further, since the sample was used to s two empirical problems (finding the "true" algebraic forms of struct equations as well as estimating parameters of these equations), would feel more comfortable if corroborative evidence were avail; Finally, the short time span involved in that sample did not pe inference regarding stability of scale economies over time.

Two new samples were used in a second empirical study. The K\ sample [23] covered twenty-four years (1939–56, 1958–63) but its ol vations were averages of data from 53 to 100 relatively small f grouped into seven to nine circulation intervals in each year. The S] sample was stratified by circulation intervals and was drawn from [24]. While its observations were individual rather than grouped represented large as well as small firms, only prices of national ad tising and total advertising output could be observed. Since nei sample permitted observation of all variables used in the first st: substantial model modification was necessary.

If all subscribers were identical in the eyes of advertisers and if

vertisers were willing to pay a certain price per inch of space in a news-
paper of given circulation, they ought to be willing to pay twice as much
per inch if circulation doubles. Alternatively, they should be willing to
pay the same price per inch per subscriber.

This suggests a commonly used scalar measure of advertising called
"subscriber-inches." This measure is simply column inches of published
advertising space multiplied by the circulation of the issue in which it
appears.[8]

Subscriber-inch price quotations are available and are known to de-
cline with paid circulation among newspaper firms [5]. By implicitly
assuming that observed prices are equal to per unit costs and by ex-
plicitly assuming that each cost-size observation deviates randomly
from a long-run cost function common to all firms, Ferguson inferred
the existence of scale economies in production of subscriber-inches of
advertising from this evidence [5, appendix 2].

Using different data, retaining Ferguson's second assumption, but
replacing his first with the assumption that each firm regards advertising
demands as parametrically given and chooses profit maximizing prices
accordingly, the study reported here reached the same conclusion. It
also found that subscriber-inch scale economies have remained stable
since 1939 and that they extend to all observable scales of output.

Note that declining average cost of advertising space or of circulation
copies implies declining cost per subscriber-inch of advertising, but that
the converse implication need not be valid. Hence, the evidence of this
study supports but does not imply findings of the first study.

The Model and Results

The samples permitted separating demands for space by national,
local-retail, and classified advertisers. Elasticities of demand for column
inches of space with respect to prices charged per subscriber-inch were
assumed constant for each kind of advertising.[9] Multiplying the implied
demand equations by daily circulation, we obtain

$$z_{ikt} = \alpha_{ikt} y_{kt} p_{ikt}^{\beta_{it}} u_{ikt} \quad i = 1, 2, 3.$$

z_{ikt} is the quantity of subscriber-inches demanded from and p_{ikt} is the
price per subscriber-inch charged by the kth firm in the tth time period
for the ith kind of advertising (national, local, or classified), and y_{kt} is

---

[8] A standard unit in the industry is the "milline" which is based on an agate line rather than
a column inch measure of space. A subscriber-inch is equal to fourteen millionths of a milline.

[9] As indicated, this assumption would be realistic if advertisers regarded all actual and
potential subscribers to a given newspaper as identical. Conclusions reported here do not
depend critically on the assumption of constant elasticity coefficients, but identification of
regression coefficients as cost parameters does depend on the assumed algebraic form of the
demand functions.

the firm's circulation. u_{ikt} is a random variable, α_{ikt} is a constant vary-ing among firms to represent changes in unobserved exogenous variables, and the elasticity coefficient β_{it} may vary over time and among kinds of advertising but is constant and identical for all firms.

As in [3], it was necessary to assume that subscriber demand is inde-pendent of the quantity and kind of advertising published. Character-ization of subscriber demand is not needed.

It was assumed that long-run costs are determined by

$$C_{kt} = \gamma_t z_{kt}^{\delta t} v_{kt} + c_{kt}$$

where C_{kt} is annual cost for and z_{kt} is annual combined advertising subscriber-inch output by the kth firm in the tth year. v_{kt} is a random variable and c_{kt} represents that part of costs which do not vary with subscriber-inch output but may vary across firms, over time, and with other dimensions of output. γ_t and δ_t are constant and identical for all firms but may change over time.[10]

$(\delta_t - 1)$ may be interpreted as the elasticity with respect to subscriber-inch output of that part of average cost which varies with subscriber-inch output. Provided $0 \leq c_{kt}$, a finding that $(\delta_t - 1) < 0$ implies the existence of scale economies in subscriber-inch production.

With specification of a circulation demand function, these assump-tions would permit construction of an eight equation model jointly de-termining profit maximizing (hence observable) prices and outputs. Because exogenous variables needed for statistical identification could not be observed, the model is empirically useless in this form. However, the arbitrary assumptions of separability in cost and demand permit derivation of equations relating profit maximizing subscriber-inch prices and outputs.

Differentiating the implied profit function with respect to subscriber-inch prices, we have

$$(\beta_{it} + 1)z_{ikt} - (\gamma_t \delta_t z_{kt}^{\delta t - 1} v_{kt})(\beta_{it} z_{ikt}/p_{ikt})w_{ikt} = 0 \quad i = 1, 2, 3$$

where the multiplicative random variables w_{ikt} are introduced to permit disturbance in these necessary conditions of profit maximization. If prices and outputs are nonzero and provided $\beta_{it} \neq -1$, these reduce to

$$p_{ikt} = \frac{\beta_{it}\gamma_t\delta_t}{\beta_{it}+1} z_{kt}^{\delta t - 1}(v_{kt}w_{ikt}) \quad i = 1, 2, 3$$

which is linear with constant coefficients under a logarithmic trans-

---

[10] [8] and [11] examine properties of additively separable functions such as the assumed cost function. Implicit in this construction is the assumption that differences in price among kinds of advertising are due to price discriminatory behavior alone. The results are not entirely consistent with this assumption.

formation. A least squares estimate of the elasticity coefficient $(\delta_t - 1)$ should provide evidence of scale economies, but it must be interpreted with caution since the compound disturbance $(v_{kt}w_{ikt})$ may well be correlated with the chosen "independent" variable.

Results using the Kuhns sample are summarized in Table 2. Estimates of $(\delta_t - 1)$ were on the order of $-.3$ indicating that average cost variable with subscriber-inch advertising output declines about 3 percent for each 10 percent increase in subscriber-inch output.

TABLE 2

LOGARITHMS OF SUBSCRIBER-INCH PRICES REGRESSED ON THE LOGARITHM
OF TOTAL SUBSCRIBER-INCH OUTPUT; SUMMARY FOR
TWENTY-FOUR YEARS USING THE KUHNS SAMPLE

	National Advertising Prices	Local-Retail Advertising Prices	Classified Advertising Prices
Mean regression coefficient	−.282	−.343	−.356
Standard deviation from the mean	.019	.026	.031
Mean standard error of the coefficients	.019	.023	.024
Mean constant term	2.928	3.041	3.186
Standard deviation from the mean	.146	.162	.135
Mean standard error of constant terms	.068	.084	.088
Mean corrected R^2	.969	.967	.961
Standard deviation from the mean	.014	.013	.015
Mean sample size	8.4	8.4	8.4

As a weak test for estimator bias, coefficients were also estimated with the roles of "independent" and "dependent" variables reversed. These elasticity estimates were smaller, but the greatest and average differences were $-.023$ and $-.010$, respectively. While quantification of bias is not possible, the stable estimates and consistently good fits argue against rejection of a scale economy hypothesis.

The coefficient estimates showed no discernible time trend with graphical and regression analysis. Since any bias present should be time stable, the evidence indicates that scale economies in advertising subscriber-inch output have not changed since 1939.[11] A slight decline in the estimates during the later years of World War II suggests that newsprint rationing may have favored larger firms.

The method of paired variates was used to test for differences in coefficient estimates among the three kinds of advertising. A possible explanation of the significantly larger national advertising elasticities may be found in the International Typographical Union's "bogus" rule. National ads are frequently printed using advertiser prepared stereo-

[11] Unreported independent but scarcely conclusive evidence suggests that scale economies have been stable since 1923 at least.

type mats while local and classified ads, whose elasticities do not differ significantly, are nearly always type-set in the firm's own composing-room. The bogus rule requires that advertiser prepared mats be reset in the publishing firm's shop whenever straight-time compositors are available. Larger firms more frequently have union contracts and available straight-time compositors, so the bogus rule is biased against a finding of scale economies in national advertising while local and classified advertising are not affected.

The model was also used with the SRDS sample where the only observed price was that of national advertising. This resulted in an elasticity estimate of $-.269$ and a constant term of 2.277 with standard errors of .015 and .092, respectively. R^2 was .886 for 45 observations.

As expected, the equation did not fit the individual firm observations of this sample as well as the grouped observations of the Kuhns sample, but the fit was good enough to indicate that previous results were not an accident of the aggregation process used to construct the Kuhns sample.

Apparently results from the Kuhns sample remain valid when a full range of newspaper sizes is considered. The new value of the elasticity coefficient $(-.269)$ falls within one standard deviation (.019) of the mean value obtained from the Kuhns sample $(-.282)$.

Conclusion

The results of these two studies are persuasive evidence that scale economies are an important if not the most important determinant of the isolated structure of the daily newspaper industry, but they have obvious limitations. Most of the empirical study was limited to fairly small firms and, even there, the terms of the trade-off between desirability and feasibility in model building became quite painful. No study was made of multiple publication establishments. Further, many would object to the use of profit maximizing behavior as a maintained hypothesis.

I hope that this preliminary effort will encourage other investigators to remedy its flaws and to extend its results to a broader sample of newspaper firms. Further evidence regarding changes over time is particularly needed.

Application of the indirect methods of cost parameter estimation to other industries may be possible. The directions suggested by Mundlak [13] for similar estimation in the context of production functions should also be explored.

In the absence of a well-developed theory of demand for variable or differentiated products, I have been forced to use such imprecise terms as "tailor-made" and "common-denominator." I would like to conclude

with a plea for the development of a demand theory counterpart to Chamberlin's theory of monopolistic competition.

REFERENCES

1. E. H. Chamberlin, *The Theory of Monopolistic Competition* (Harvard Univ. Press, 1960).
2. ———, "The Product as an Economic Variable," *Q.J.E.*, Feb., 1953, pp. 1–29.
3. W. M. Corden, "The Maximisation of Profit by a Newspaper," *Rev. of Econ. Studies*, 1953, pp. 181–90.
4. G. Debreu, *Theory of Value*, Cowles Monograph 17 (Wiley, 1959).
5. J. M. Ferguson, *The Advertising Rate Structure in the Daily Newspaper Industry* (Prentice-Hall, 1963).
6. F. M. Fisher, "Identifiability Criteria in Nonlinear Systems: A Further Note," *Econometrica*, 1965, pp. 197–205.
7. M. Friedman, "Comment," *Conference on Business Concentration and Price Policy* (Princeton, 1955).
8. S. M. Goldman and H. Uzawa, "A Note on Separability in Demand Analysis," *Econometrica*, 1964, pp. 387–98.
9. N. Kaldor, *Essays on Value and Distribution* (Free Press, 1960).
10. L. R. Klein, *A Textbook of Econometrics* (Evanston, 1953).
11. W. W. Leontief, "A Note on the Interrelation of Subsets of Independent Variables of a Continuous Function with Continuous First Derivatives," *Bull. of the Amer. Math. Soc.*, 1947, pp. 343–50.
12. J. W. Markham, "Economic Analysis of Competition in the Daily Newspaper Business," *Newspapers, 1963*, Presentation by the American Newspaper Publishers Association before the Antitrust Subcommittee of the House Judiciary Committee, an ANPA Publication, N. Y.
13. Y. Mundlak, "Transcendental Multiproduct Production Functions," *Int. Econ. Rev.*, 1964, pp. 273–84.
14. M. Olson and D. McFarland, "The Restoration of Pure Monopoly and the Concept of Industry," *Q.J.E.*, Nov., 1962.
15. R. H. Ray, "Competition of the Newspaper Industry," *J. of Marketing*, 1951, pp. 444–56.
16. W. B. Reddaway, "The Economics of Newspapers," *Econ. J.*, 1963, pp. 201–18.
17. P. Samuelson, "The Pure Theory of Public Expenditure," *Rev. of Econ. and Statis.*, 1954, pp. 387–89.
18. H. Uzawa, "Duality Principles in the Theory of Cost and Production," *Int. Econ. Rev.*, 1964, pp. 216–20.
19. A. A. Walters, "Production and Cost Functions: An Econometric Survey," *Econometrica*, 1963, pp. 1–66.
20. Audit Bureau of Circulations, current "Audit Reports" for Midwestern Dailies, Chicago, Illinois, 1964.
21. Editor and Publisher Magazine, "Annual Linage Reports," first Apr. issue, 1959, 1961, 1963, and 1965.
22. ———, *Yearbooks* for 1923, 1958, 1960, 1962, 1963, 1964.
23. Newspaper Audit and Research Bureau, Division of M. S. Kuhns and Company, "Special Statistical Report Showing Group Averages Taken from Groups of Non-Metropolitan Newspapers," 1939–56, 1958–63, Chicago, Illinois (24 annual issues).
24. Standard Rate and Data Service, Inc., "Newspaper Rates and Data," Jan. issue, 1959, 1961, 1963, 1965 (Skokie, Illinois).

RATIONAL CHOICE AND PATTERNS OF GROWTH IN A MONETARY ECONOMY*

By MIGUEL SIDRAUSKI

Massachusetts Institute of Technology

Concluding his excellent survey of recent monetary theory, Harry Johnson [4] suggested that future developments in this field should come from attempts "to break monetary theory loose from the mould of short-run equilibrium analyses, conducted in abstraction from the process of economic growth and accumulation, and to integrate it with the rapidly developing theoretical literature on economic growth."

This paper summarizes an attempt to deal with these issues. Like most theoretical work in rapidly growing fields, it is incomplete and the assumptions on which it is based are relatively crude abstractions. These abstractions, however, allow us to explore certain aspects of the interaction of the real and the monetary phenomena in a model of economic growth in which money, being government noninterest bearing debt, is introduced as an alternative asset to real capital.

Most of the recent work in this field[1] has centered on the analysis of the patterns of growth of a monetary economy by postulating alternative plausible saving functions and demand functions for money. What differentiates this product is the fact that, in line with Patinkin's [6] presentation of the neoclassical theory of money, and with the classical Fisherian theory of saving [2], it is based on an explicit analysis of individuals' saving behavior, viewed as a process of wealth accumulation aimed at maximizing some intertemporal utility function.

The first part of the paper describes the representative economic unit of an idealized economy and it analyzes the constraints imposed on its maximizing behavior. Section II is concerned with the optimizing conditions and presents the derivation of the demand functions for consumption, cash balances, and the stock of capital. The third section introduces an expectations formation hypothesis and presents a simple aggregative macroeconomic model in which the demand functions for assets and consumption are those which were derived from the analysis of the maximizing behavior of individual economic units. The final part of the paper considers the short-run and long-run effects of a change in the rate of monetary expansion as well as the stability of the equilibrium growth path in a monetary economy.

[1] What I have in mind here is the work by Tobin [9], Gurley and Shaw [3], Johnson [5], and Sidrauski [7].

I. *The Model*

The basic economic unit in our model is the representative family. Its welfare at any point in time is measured by a time invariant utility function of the form

$$(1) \qquad\qquad U_t = U(c_t, z_t)$$

where c_t stands for the flow of real consumption per unit of time, and z_t for the flow of services per unit of time derived from holdings of real cash balances, both variables being expressed in per capita terms. To simplify, we will assume that the flow of services derived from the holdings of real cash balances is proportional to the stock and, by an appropriate choice of units, we make the factor of proportionality equal to one.

$$(2) \qquad\qquad z_t = m_t = M_t/p_t N_t$$

M_t represents the holdings of nominal cash balances by the economic unit, N_t the number of individuals in the economic unit and p_t the money price of the only commodity produced in our model. The instantaneous utility function can then be written

$$(3) \qquad\qquad U_t = U(c_t, m_t)$$

It is assumed that the utility function is strictly concave with continuous first and second derivatives[2] and that both commodities are not inferior.[3] We further assume that the total welfare (W) associated with any particular time path (c_t, m_t) can be represented by the utility functional

$$(4) \qquad\qquad W = \int_0^\infty [U(c_t, m_t)] e^{-\delta t} dt$$

$\delta > 0$ being the subjective rate of time preference of this family.

At each moment of time the behavior of the economic unit is subject to two constraints, one in terms of stocks and the other one in terms of flows. The stock constraint requires that the total endowment of real nonhuman wealth (a_t), be allocated between capital (k_t) and real cash balances in such a fashion that[4]

$$(5) \qquad\qquad a_t = k_t + m_t$$

On the other hand, the flow constraint requires that at any time t disposable income has to be equal to consumption plus saving. Assuming that the production function is linear homogeneous, the capital stock produces an amount $y(k_t)$ of homogeneous output.[5] If we add to this

---

[2] This condition implies that $U_{cc} < 0$, $U_{mm} < 0$, and $J = U_{cc}U_{mm} - U_{mc}^2 > 0$.

[3] This requires: $J_1 = U_{mm} - U_{cm}U_m/U_c < 0$ and $J_2 = U_{cc}U_m/U_c - U_{cm} < 0$.

[4] In what follows all variables are expressed in per capita terms.

[5] We will assume that the production function is "well behaved": namely, $y(0) = 0$, $y(\infty) = \infty$, $y(0 < k < \infty) > 0$, $y'(0) = \infty$, $y'(\infty) = 0$, $y'(0 < k < \infty) > 0$, $y''(k) < 0$.

amount the real value of the net transfers that the economic unit receives from the government (v_t), we obtain the family's gross disposable income, which has to be equal to real consumption (c_t) plus gross real savings (s_t).

$$(6) \qquad\qquad y(k_t) + v_t = c_t + s_t$$

Gross real saving is the sum of gross capital accumulation (i_t) plus the gross addition to the holdings of real cash balances (x_t).

$$(7) \qquad\qquad s_t = i_t + x_t$$

Gross capital accumulation is equal to the net addition to the capital stock $(\dot{k}_t)$ plus the replacement of the depreciated capital (uk_t) plus the amount of capital accumulation required to provide the newly born members of the economic unit with the same amount of capital as the amount with which the old members are endowed (nk_t); where u is the instantaneous rate of depreciation of capital and n is the instantaneous rate of growth of the number of individuals in the family.

$$(8) \qquad\qquad i_t = \dot{k}_t + (u + n)k_t$$

Similarly, the gross accumulation of real cash balances is equal to

$$(9) \qquad\qquad x_t = \dot{m}_t + (\pi_t + n)m_t$$

where π_t is the expected rate of change in prices.

Hence the flow constraint for this economic unit can be rewritten

$$(10) \qquad y(k_t) + v_t - (\pi_t + n)m_t - (u + n)k_t - \dot{m}_t - \dot{k}_t - c_t = 0$$

Differentiating equation (5) with respect to time and substituting into (10) we have

$$(11) \qquad \dot{a}_t = y(k_t) + v_t - (\pi_t + n)m_t - (u + n)k_t - c_t$$

Equations (5) and (11) are the stock and flow constraints under which, given the initial condition (a_0) and the values of u, n, π_t and v_t, the rational economic unit will find the time path of consumption and accumulation that maximizes the utility functional (4).

II. Maximization of the Utility Functional and Derivation of the Demand Functions

In order to solve this maximization problem we form a new function (I) such that

$$(12) \qquad I = \int_0^\infty \Big\{ U(c_t, m_t) + \lambda_t[y(k_t) + v_t - (\pi_t + n)m_t \\ - (u + n)k_t - c_t - \dot{a}_t] + q_t[a_t - k_t - m_t] \Big\} e^{-\delta t} dt$$

where λ_t is the Lagrangian multiplier attached to the flow constraint (11) and q_t the Lagrangian multiplier attached to the stock constraint (5). The conditions for a maximum are given by the Euler equations (13)–(16) together with the transversality condition (17).

(13) $U_c(c_t, m_t) = \lambda_t$

(14) $U_m(c_t, m_t) = \lambda_t(\pi_t + r_t + n)$

(15) $y'(k_t) - (u + n) = r_t$

(16) $\dfrac{\dot{\lambda}_t}{\lambda_t} = \delta - r_t$

(17) $\lim\limits_{t \to \infty} a_t \lambda_t e^{-\delta t} = 0$

where r_t is equal to the ratio of the two Lagrangian multipliers q_t/λ_t.

Equations (13)–(16) together with the constraints (5) and (11) form a system of six equations which, given the values of u, n, π_t, and v_t and given the initial stock of real wealth a_0, will describe the time path of the six endogenous variables c_t, m_t, k_t, a_t, λ_t, and r_t. The problem is therefore to find the path of these variables which, satisfying equations (5), (11) and (13)–(16) will also satisfy the transversality condition (17).[6]

From equations (13) and (14) we can solve for the quantities demanded of consumption and real cash as functions of the implicit price of consumption, λ, the implicit interest rate, r, and the expected rate of change in prices, π;

(18) $c = c^0(\lambda, r, \pi)$

(19) $m = m^0(\lambda, r, \pi)$

and from equation (15) we can solve for the quantity demanded of real capital as a function of the implicit interest rate.

(20) $k = k^0(r)$

Considering now the stock constraint (7) we determine the implicit rate of interest as a function of the stock of wealth, the implicit price of consumption and the expected rate of inflation.

(21) $a = k^0(r) + m^0(\lambda, r, \pi)$

namely

(22) $r = r(a, \lambda, \pi)$

and substituting back into (18), (19) and (20) we can write the demand for consumption, capital and real cash as functions of the stock of real

---

[6] In what follows the time subscripts will be used only where necessary for a better understanding of the text.

nonhuman wealth, the implicit price of consumption and the expected rate of change in prices

(23) $c = c'(a, \lambda, \pi)$

(24) $m = m'(a, \lambda, \pi)$

(25) $k = k'(a, \lambda, \pi)$

Finally, given the expected rate of inflation and government transfer payments, the pair of differential equations (11) and (16) determine the time path of the implicit price of consumption and the stock of wealth. The laws of motion of our system are shown in the phase diagram below. To verify that (λ^*, a^*) is a saddle point we solve the characteristic equation for the linear Taylor approximation to (11) and (16). For the characteristic roots to be real and opposite in sign, guaranteeing that (a^*, λ^*) is a local saddle point the following condition must hold[7]

(26) $(\pi + n)J_1 + J_2 < 0$

This condition will be satisfied for any expected rate of change in prices which is not smaller than the rate of growth of population. It should be noted that although the slope of the $\dot{a} = 0$ schedule in the phase diagrams may be positive or negative, the solution (a^*, λ^*) is a local saddle point provided that (26) holds.

Therefore, given the initial holdings of assets a_0, there is only one time path of a and λ which will satisfy the Euler conditions (13)–(16), the constraints (5) and (11) and the transversality condition (17). This path
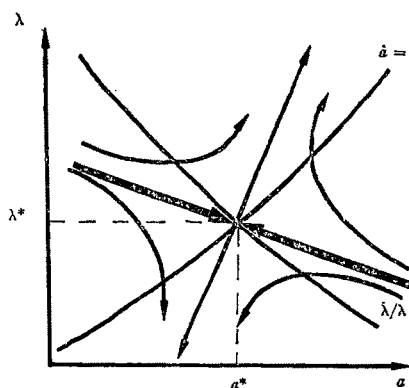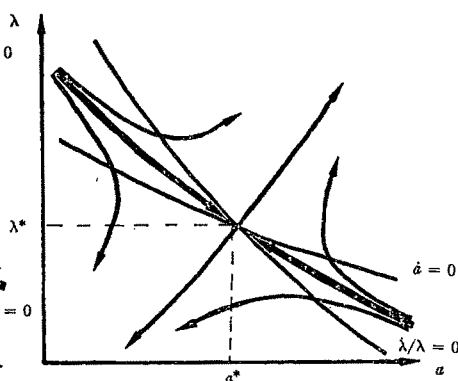

FIGURE 1a   FIGURE 1b

is indicated by the heavy arrows in Figures 1a and 1b. All other paths (light arrows) fail to satisfy condition (17). For each total stock of wealth there is one implicit price of consumption λ which will determine

---

[7] This condition is also derived in [8].

the optimum allocation of the stock of wealth between capital and real money and the optimum allocation of the income flow between consumption and net saving. This implicit price is the value of λ on the heavy arrows that corresponds to the given level of wealth. Clearly, a change in the expected rate of inflation as well as a change in the real value of government transfers will shift the optimum path, hence

$$(27) \qquad\qquad \lambda = \lambda(a, \pi, v)$$

By substituting (27) into (23), (24) and (25) we have shown that the quantities of capital, real money and consumption demanded are functions of total wealth, the expected rate of inflation and the net government transfers to the private sector.

$$(28) \qquad\qquad c = c(a, \pi, v)$$

$$(29) \qquad\qquad m = m(a, \pi, v)$$

$$(30) \qquad\qquad k = k(a, \pi, v)$$

III. *The Expectations Hypothesis and the Macroeconomic Model*

It will be assumed that expectations are induced, i.e., that individuals take past rates of change in prices into account in forming their expectations about a "normal rate of change in prices," on the basis of which they determine the amounts of real cash balances, capital and consumption that they demand at each moment of time. In particular, our hypothesis will imply that when individuals realize that their expectations did not materialize they partially revise what they consider to be the "normal rate of change in prices." This is the so-called "adaptive expectations" model that was originally introduced by Cagan [1] and which can be expressed in terms of the following differential equation

$$(31) \qquad\qquad \dot{\pi} = b(\dot{p}/p - \pi); \quad b > 0$$

From the point of view of the economy as a whole, market equilibrium requires that at each moment of time the demand for money be equal to the total money supply. Thus, given the stock of money M, the equilibrium condition in the money market can be written as follows

$$(32) \qquad\qquad \frac{M}{pN} = m(a, \pi, v)$$

where

$$(33) \qquad\qquad a = \frac{M}{pN} + k$$

The government levies taxes and makes transfer payments to the

private sector but it does not undertake any public expenditures. We assume that the excess of transfer payments over taxes is entirely financed by the creation of government noninterest bearing debt which we call money. Therefore, the excess of transfers over taxes is exactly equal to the amount of money issued per unit of time. We also assume that each economic unit in the economy receives the same amount of net transfers; namely, that taxes and transfers are of the per capita type. Hence we can write

$$(34) \qquad v = \frac{\dot{M}}{Np} = \theta m$$

where $\theta = \dot{M}/M$. Equations (31)–(34) form a system of four equations in seven unknowns, M, N, k, p, a, v and π. In order to have a complete system that describes the behavior of the economy through time, three additional equations are required. To complete the system we assume that the government maintains a constant rate of monetary expansion and population grows at a constant rate

$$(35) \qquad \dot{M}/M = \theta_0$$

$$(36) \qquad \dot{N}/N = n$$

Finally, since all output that is not consumed is necessarily used for capital accumulation, the rate of change in the capital stock is given by the following expression

$$(37) \qquad \dot{k} = y(k) - (u + n)k - c(a, \pi, v)$$

We now have a system of seven equations in seven unknowns that describes the time path of our simplified economy. However, before we provide such a description it should be clear from the system of equations that a major difference arises when we go from the analysis of the individual economic unit to the analysis of the economy as a whole. While for the individual economic unit real wealth, the real value of the net government transfers and the expected rate of inflation are the variables that are given at each moment of time, for the economy as a whole these variables are the stock of capital, total population, the nominal quantity of money and the rate at which the government is increasing this quantity. Therefore, considering equations (32)–(34) we can write

$$(38) \qquad \frac{M}{pN} = \hat{m}\,(k, \theta, \pi)$$

Thus, given M, N, k, θ, and π the price level is determined and therefore the real value of privately held cash is also determined.

Substituting equations (33), (34) and (38) into (37) the rate of change in the capital stock can then be written as

$$(39) \qquad \dot{k} = y(k) - (u + n)k - \hat{c}(k, \theta, \pi)$$

where $\hat{c}$ is the consumption level desired for the price level p that equilibrates the money market at each moment of time.

The system of equations that describes the behavior of the economy through time is now given by (31), (35), (36), (38) and (39), the variables of the system being N, M, k, π and p. Differentiating equation (38) with respect to time and making use of (31), (35), (36) and (39) we express the rate of change in the expected rate of inflation as follows

$$(40) \quad \dot{\pi} = \cfrac{1}{\left[1 + b \, \dfrac{\partial \hat{m}}{\partial \pi} \, \dfrac{1}{\hat{m}} \right]}$$

$$\cdot \left\{ \theta - \pi - n - \left[y(k) - (u + n)k - \hat{c}(k, \theta, \pi) \right] \right\}$$

Given the rate of monetary expansion the system of differential equations (39), (40) describes the time path of the economy.

By setting $\dot{\pi}$ and $\dot{k}$ equal to zero in equations (39) and (40) it follows that along an equilibrium growth path consumption is equal to net output and the expected rate of inflation is equal to the difference between the rate of monetary expansion and the economy's rate of growth, which along such a path is equal to the rate of population growth. Hence,

$$(41) \qquad c^* = y(k^*) - (u + n)k^*$$

$$(42) \qquad \pi^* = \theta - n$$

IV. Short-run and Long-run Effects of a Change in the Rate of Monetary Expansion and the Stability of Equilibrium Growth

A question that naturally comes to mind is what are the conditions under which the equilibrium growth path characterized by equations (41) and (42) is stable? In what follows we will consider the local stability of our model and for this purpose we have to indicate first how changes in the capital stock, the expected rate of inflation and the rate of monetary expansion affect the demand for consumption and for real cash balances. It can be shown that, in the neighborhood of the equilibrium growth path, consumption and the real value of the stock of money are increasing functions of the stock of capital and of the rate of monetary expansion and decreasing functions of the expected rate of change in prices, namely

$$(43) \qquad \frac{\partial \hat{c}^*}{\partial k^*} > 0 \qquad \frac{\partial \hat{c}^*}{\partial \theta} > 0 \qquad \frac{\partial \hat{c}^*}{\partial \pi^*} < 0$$

$$(44) \qquad \frac{\partial \hat{m}^*}{\partial k^*} > 0 \qquad \frac{\partial \hat{m}^*}{\partial \theta} > 0 \qquad \frac{\partial \hat{m}^*}{\partial \pi^*} < 0$$

An increase in the capital stock raises consumption and the real value of the stock of cash of two accounts. First, because it raises disposable income and, second, because it is associated with an increase in the stock of real wealth. In addition, the increase in the stock of capital lowers its marginal product and therefore results in a decrease of the cost of holding cash, further raising the demand for the alternative asset. Net government transfers to the private sector are assumed to be financed entirely by money creation and therefore an increase in the rate of monetary expansion is equivalent to a rise in private disposable income due to higher government net transfers. The increase in real income raises the demand for both consumption and real cash balances. Finally, an increase in the expected rate of change in prices is equivalent to a rise in the rate of depreciation of one of the assets and it is therefore associated with a decrease in private disposable income which, in turn, lowers the demand for consumption and real money. In addition, the increase in the expected rate of inflation raises the opportunity cost of holding cash, thereby further reducing the demand for this asset.

With this information we now verify what are the conditions under which (k^*, π^*) is a stable solution to the pair of differential equations (39) and (40). For this purpose we solve the characteristic equations for the Taylor approximation to (39) and (40) at (k^*, π^*). From this exercise it follows that the necessary and sufficient conditions for the stability of the equilibrium growth path are

$$(45) \qquad y'(k^*) - (u + n) - \frac{\partial \hat{c}^*}{\partial k^*} < 0$$

$$(46) \qquad \left(1 + \frac{\partial \hat{m}^*}{\partial \pi^*} \frac{1}{\hat{m}^*} b \right) > 0$$

The condition in (45) reflects the fact that an increase in the capital stock has two opposite effects on the rate of capital accumulation. On the one hand, it increases real consumption, thereby lowering the rate of change of the capital stock, and, on the other hand, it raises net output and therefore stimulates capital accumulation. Stability in this model requires the rate of capital accumulation to be a decreasing function of the capital stock.

Since an increase in the expected rate of inflation reduces the demand

for cash, it therefore results in a rise in prices. The increase in prices raises people's expected rate of inflation, further reducing the demand for money and causing a new rise in prices. Given that we assume that there is no lag in the adjustment of the actual to the desired stock of cash, the stability of the system depends on the existence as well as the magnitude of the expectations lag, as indicated by the expression in (46).

If the condition in (46) is satisfied, the solution (k^*, π^*) to the pair of differential equations (39) and (40), as indicated by the phase diagram in Figure 2, is a stable node. Provided the system is stable, a constant rate of monetary expansion will therefore guarantee a monotonic approach to the equilibrium growth path.
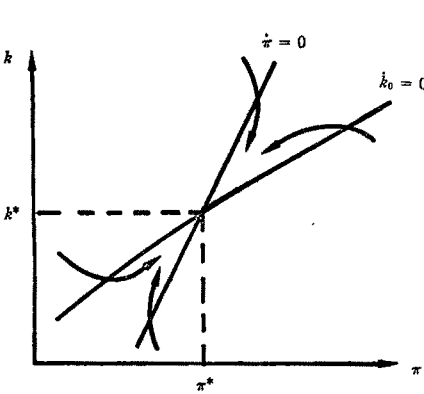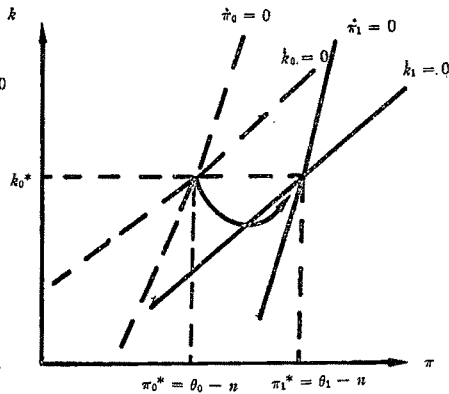

FIGURE 2   FIGURE 3

Consider now a situation in which the economy has reached its equilibrium growth path with $k^* = k_0^*$ and $\pi^* = \pi_0^*$ for $\theta = \theta_0 = \pi_0^* - n$. Suppose now that the government decides to change the rate of monetary expansion from θ_0 to θ_1, where $\theta_1 > \theta_0$. The first impact of this change is an increase in consumption which lowers capital accumulation, as well as an increase in the rate of change in prices which raises the rate of change of the expected rate of inflation. Both the $\dot{k} = 0$ and $\dot{\pi} = 0$ schedules shift to the right (Figure 3). Since we already know that the desired long-run stock of capital is determined only by its rate of depreciation, the rate of population growth and the subjective rate of time preference,[8] and, given that none of these variables are affected by the change in θ, the $\dot{k} = 0$ and the $\dot{\pi} = 0$ schedules will intersect at the same $k^* = k_0^*$ and at π_1^*, where $\pi_1^* = \theta_1 - n > \pi_0^*$. It therefore follows that while a change in the rate of monetary expansion lowers the short-

[8] This can be proved by setting equation (16) equal to zero and substituting into (15). The result of this operation is $y'(k^*) - u = (\delta + n)$.

run rate of capital accumulation, it does not affect the economy's long-run stock of capital.

V. *Conclusion*

Under the assumptions set out in Section I we have proved that in a growth model in which utility maximizing families are the basic economic unit of the system, the long-run capital stock of the economy is independent of the rate of monetary expansion. A rise in the rate of monetary expansion results in an equal absolute increase in the rate of change in prices; it reduces the stock of real cash but it does not affect steady state consumption. It therefore follows that the higher the rate of monetary expansion the lower will be the steady state level of utility. In the short run, an increase in the rate of monetary expansion is equivalent to a rise in government transfers to the private sector. It therefore results in an increase in consumption and a fall in the rate of capital accumulation.

Finally, we have also shown that in the absence of perfect foresight, utility maximization guarantees the fulfillment of only one of the steady state stability conditions; namely, that the rate of capital accumulation be a decreasing function of the capital stock. Since there is no lag in adjustment of the actual to the desired stock of cash, an additional, necessary but not sufficient, condition for the stability of the steady state is the existence of a lag in the formation of expectations.

REFERENCES

1. P. Cagan, "The Monetary Dynamics of Hyperinflation," in *Studies in the Quantity Theory of Money*, ed., M. Friedman (Univ. of Chicago Press, 1956).
2. I. Fisher, *The Theory of Interest* (New York, 1930).
3. John G. Gurley and Edward S. Shaw, *Money in a Theory of Finance*, with a mathematical appendix by A. C. Enthoven (Washington, 1960).
4. Harry G. Johnson, "Monetary Theory and Policy," *A.E.R.*, June, 1962.
5. ———, "The neo-classical One Sector Growth Model, A Geometrical Exposition and Extension to a Monetary Economy," *Economica*, Aug., 1966.
6. D. Patinkin, *Money, Interest and Prices* (Row-Peterson, 1956).
7. M. Sidrauski, "Growth in a Monetary Economy" (Jan., 1966, mimeo.).
8. ———, "Inflation, Optimum Consumption and Real Cash Balances," a paper presented at the meetings of the Econometric Society, New York, Dec., 1965.
9. J. Tobin, "Money and Economic Growth," *Econometrica*, Oct., 1965.

THE CANADIAN EXPERIENCE WITH FLEXIBLE EXCHANGE RATES[*]

By JOHN PIPPENGER
University of California, Santa Barbara

A foreign exchange market operating under a system of flexible rates may be classified, along with stock markets and commodity futures markets, as a speculative market. Recent studies of stock and commodity futures markets based on an analysis of price changes have resulted in an improved understanding of these markets.[1] This paper applies analysis of price changes to the Canadian experience with flexible exchange rates. Autocorrelation and spectral analysis of percentage changes in the daily U.S. price of the Canadian dollar indicate the presence of a stabilizing influence in the market—an influence that is identified as intervention by the Canadian Exchange Fund.

On September 30, 1950, the Canadian government announced the decision to free the exchange rate. Although the price of the Canadian dollar was no longer fixed, the foreign exchange market was not a completely free market. The demand for foreign exchange was still limited to those transactions permitted under the Foreign Exchange Control Act and Regulations. During 1951 there was a gradual reduction in exchange controls, and on December 14, 1951, all foreign exchange restrictions were suspended. This suspension became permanent when the Exchange Control Act was later repealed. From January, 1952, to November, 1960, the Canadian foreign exchange market was a free market subject only to official intervention designed to eliminate what were believed to be unwarranted short-run fluctuations.

Part I establishes a definition of stabilizing and destabilizing activity and considers the type of exchange rate behavior implied by an attempt on the part of an exchange fund to stabilize the rate. Part II presents the results of autocorrelation and spectral analysis of daily changes in the U.S. price of Canadian dollars from January, 1952, to November, 1960. Part III presents the conclusion that the Canadian Exchange Fund stabilized the exchange rate by reducing short-run fluctuations without materially affecting long-run movements.[2]

[1] Paul H. Cootner (ed.), *The Random Character of Stock Market Prices* (M.I.T. Press, 1964), contains an excellent collection of articles on the behavior of stock prices, and an extensive bibliography on price behavior in both stock and commodity futures markets can be obtained from the references cited in the various articles.

[2] For a different approach, see Harry C. Eastman and Stefan Stykolt, "Exchange Stabilization in Canada, 1950-4," *Canadian J. of Econ. and Polit. Sci.*, May, 1956, pp. 221-33.

I. *Definition and Measurement*

Definitions of stabilizing and destabilizing activity have generally taken one of two forms. Under the first approach, an activity, usually speculation, is defined as destabilizing (stabilizing) when it moves the actual rate away from (toward) the equilibrium exchange rate. This definition appears to be empirically empty. Equilibrium exchange rates exist in the minds of economists; they do not exist in the market place. There are as many equilibrium exchange rates as there are possible theories designed to explain exchange rates. The arbitrary selection of one of these equilibria as *the* equilibrium does not appear to be a satisfactory solution to the problem.

Under the second approach, an activity is destabilizing (stabilizing) when it increases (decreases) the size of price fluctuations. For the purposes of this paper, an activity is defined as being destabilizing (stabilizing) when it increases (decreases) the variance in the percentage changes in the exchange rate.

If a given market could be observed first with and then without the activity being analyzed, this definition could be applied in a straightforward manner. However, in most situations only the exchange rate containing the influence of the activity being analyzed is observed. If the rate in the absence of this influence is unknown, how can it be determined whether the observed variance in the percentage changes in the rate is larger or smaller than it would have been had the influence been absent? Although it is impossible to determine the effect of a given activity directly, it is possible to identify its effect indirectly.

Diagram A in Figure 1 shows the market demand for and supply of Canadian dollars as a function of the U.S. price of the Canadian dollar. *SD* in Diagram B of Figure 1 is the market demand for Canadian dollars minus the market supply of those dollars. *EE* in Diagram B represents one possible strategy for an exchange fund committed to ironing out short-run fluctuations in the exchange rate without resisting long-run movements. Under this strategy, the fund sells domestic currency as the rate rises above the rate in the previous period and buys as the rate falls below the previous rate.[3]

The model may be described as follows. At noon each day all of the participants in the exchange market meet and each participant, in-

---

[3] *SD* in Figure 1 may be written as $\$_c/t = a_t + bR_t$, where R_t is the U.S. price of the Canadian dollar in period t, $b < 0$, $a_t > 0$, and $\$_c/t$ measures Canadian dollars per unit of time. *EE* in Figure 1 is given by $\$_c/t = e(R_t - R_{t-1})$, $e > 0$. Solving for R_t yields

$$R_t = \frac{-a_t - eR_{t-1}}{b - e}.$$

The patterns displayed by the spectrum and autocorrelation function are not substantially affected if R_{t-1} is replaced by $\bar{R}$, a "normal" rate obtained by taking a declining weighted average of past rates, the sum of the weights being one.
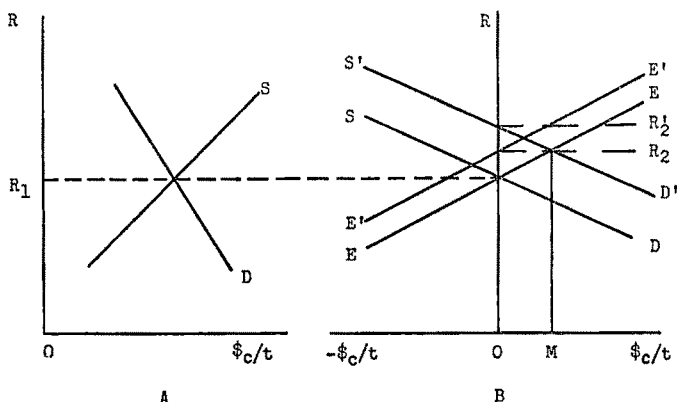
FIGURE 1
DEMAND FOR AND SUPPLY OF
CANADIAN DOLLARS

cluding the exchange fund, specifies how many dollars he is willing to buy or sell at each alternative exchange rate, the relationship between price and quantity being linear for each participant. This information is put into a computer which determines the rate that clears the market.

Some may object to using such a simple and unrealistic model to analyze the exchange market and the effect of an exchange fund's actions on the behavior of exchange rates. I suggest that it is up to those who make such an objection to show that a more realistic model (e.g., a model that generates exchange rates as a continuous function of time) which is consistent with observed exchange rate behavior, yields different conclusions concerning the effect of an exchange fund's actions on day-to-day movements in the rate.

Let the excess demand function for period one be SD in Diagram B. In period two, the excess demand function shifts to $S'D'$. (EE in period two intersects the vertical axis at R_1 because the rate in period one is R_1.) In the absence of the fund, the rate would move to R_2', but the fund sells OM Canadian dollars as the U.S. price of the Canadian dollar rises and, as a result, the rate rises only to R_2. Because of the movement in the rate to R_2, EE shifts to $E'E'$ in period three. If the shift from SD to $S'D'$ is temporary and the excess demand function moves back to SD in period three, then the fund's actions have reduced the size of the fluctuations. If the shift is permanent, then the rate moves by a series of steps toward R_2'.

This strategy reduces short-run fluctuations in the rate without preventing long-run movements and, by slowing down the speed with

which the rate adjusts to new long-run situations, introduces a tendency for sustained movements. Although sustained movements have generally been associated with destabilizing activity, they appear more likely to be the result of stabilizing activity.

If speculators sell a currency as its price rises and buy the currency as the price falls, then they also will tend to stabilize the rate. If speculators or an exchange fund follow the opposite policy, then their actions will tend to amplify exchange rate fluctuations and introduce a tendency for a rise in the rate to be followed by a fall and vice versa; i.e., a tendency for reversals.[4] Therefore, evidence of sustained movements in exchange rates supports the claim that there is a stabilizing influence in the market, while evidence of reversals supports the claim of a destabilizing influence.

Evidence from other speculative markets such as stock and commodity futures markets indicates that, as a first approximation, the price change during a given day is independent of the change during all previous days in the sense that the autocorrelation function for the precentage changes in the daily price is zero for all lags greater than zero. There is no significant tendency in stock or commodity futures prices for either sustained movements or reversals.[5]

Freely fluctuating exchange rates, however, show a small but clear tendency for sustained movements. Autocorrelation estimates for percentage changes in daily rates in freely fluctuating markets show a persistent tendency to be positive for a one-day lag, and analyses of runs show fewer than the expected number of runs given independence between the changes.[6]

Although autocorrelation and analysis of runs shed some light on the behavior of exchange rates, spectral analysis appears to be a natural tool by which to determine whether an activity has a stabilizing or destabilizing effect on exchange rates. When price changes generate a stationary time series and are independent in the sense that the autocorrelation function is zero for all lags greater than zero, then the spectrum for the series plotted against frequency will appear as a horizontal line. This indicates that each frequency is contributing the same relative amount to the variance in the series.[7]

Suppose the percentage of changes in the Canadian exchange rate in the absence of any activity by the Exchange Fund generated a stationary series in which the autocorrelations were zero for all lags

---

[4] This model of destabilizing activity can be obtained from the model described in footnote 3 by reversing the sign of e.

[5] Cootner, *op. cit.*

[6] John Pippenger, "The Behavior of Flexible Exchange Rates: Theory and Evidence" (unpublished doctoral dissertation, Univ. of California, Los Angeles, 1966), pp. 314-19.

[7] For an explanation of spectral analysis see C. W. Granger and M. Hatanaka, *Spectral Analysis of Economic Time Series* (Princeton Univ. Press, 1964), and R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra* (New York: Dover, 1958).

greater than zero. If this were so, what pattern would the autocorrelation function and spectrum for the percentage changes in the observed rate display if the Canadian Exchange Fund sold Canadian dollars as their price rose and bought Canadian dollars as their price fell?

In order to answer this question, a simulated foreign exchange market corresponding to *SD* in Diagram B of Figure 1 has been constructed so that it yields a set of exchange rates whose percentage first differences constitute a stationary time series in which the autocorrelations are zero for all lags greater than zero. The upper solid line in Figure 2 shows the spectrum for a sample of 1,000 observations from this series. (The pair of horizontal broken lines establish a 90 per-
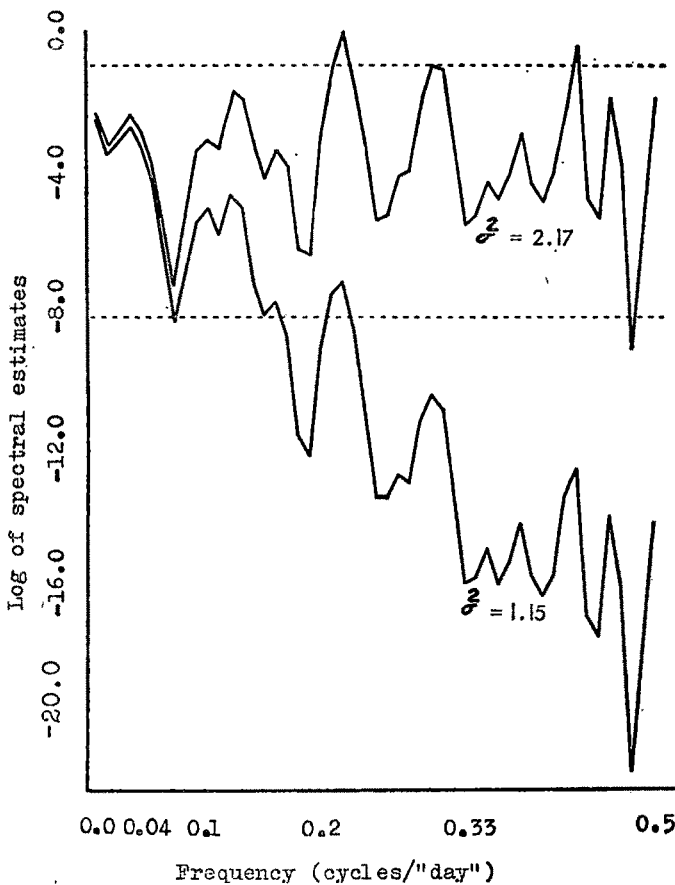

FIGURE 2
SPECTRAL ESTIMATES OF PERCENTAGE CHANGES
IN SYNTHETIC RATES
PLOTTED IN LOG
SCALE AGAINST FREQUENCY
(1,000 OBSERVATIONS)

cent "confidence" interval for the spectral estimates.) An exchange fund behaving according to *EE* in Diagram B of Figure 1 is then introduced into the simulated market. The influence of the exchange fund is increased until it reduces the variance in the percentage changes by approximately 50 percent (from 2.17 to 1.15).[8]

The autocorrelation estimates for the second series of percentage changes in the simulated rate are shown in Table 1. The effect on the autocorrelations is clear. An exchange fund that reduces the fluctuations in the rate by selling (buying) domestic currency as its price rises above (falls below) the price in the previous period introduces a tendency for the autocorrelation estimates of the percentage changes in the rate to show positive values for short lags.

The spectrum for the series analyzed in Table 1 is the lower solid line in Figure 2. The effect of the fund's activity on the exchange rate is shown by the different pattern displayed by the spectrum. In the absence of the fund, the spectral estimates trace out a more or less hori-

TABLE 1

AUTOCORRELATION ESTIMATES FOR PERCENTAGE CHANGES
IN SYNTHETIC RATE, ONE TO FIFTY PERIOD LAG
$(\sigma^2 = 1.15)$

Lag	Estimate	Lag	Estimate
1	0.327*	26	0.028
2	0.078*	27	−0.058*
3	0.029	28	−0.008
4	0.020	29	−0.021
5	0.017	30	0.031
6	0.005	31	0.005
7	0.037	32	0.024
8	0.026	33	0.009
9	0.053*	34	0.022
10	0.029	35	0.024
11	−0.044	36	−0.008
12	−0.052*	37	0.028
13	−0.002	38	0.007
14	0.007	39	0.017
15	−0.048	40	0.030
16	−0.028	41	0.027
17	−0.003	42	0.029
18	0.007	43	−0.021
19	0.022	44	−0.008
20	−0.004	45	0.008
21	−0.011	46	−0.015
22	0.012	47	−0.010
23	0.034	48	0.024
24	0.048	49	0.049
25	−0.016	50	0.022

* Significant at the 90 percent level. For the distribution of the autocorrelation estimates see R. L. Anderson, "Distribution of the Serial Correlation Coefficient," *Annals of Mathematical Statistics*, Mar., 1942, pp. 1–13.

[8] That is, *e* in footnote 3 was increased until the variance in the percentage changes in R_t was reduced by approximately one-half. For a more complete description of the models used to generate synthetic exchange rates, see Pippenger, *op. cit.*, pp. 362-71.

zontal line. Purchases and sales by the fund cause the spectral estimates for the high frequencies (short cycles) to fall while leaving the estimates for the lowest frequencies (longest cycles) almost unaffected. The exchange fund's purchases and sales in this model reduce the "day-to-day" and "week-to-week" movements in the rate without materially affecting the "year-to-year" movements.

If the exchange fund buys domestic currency as the rate rises and sells as the rate falls, then the autocorrelation estimates for the percentage changes in the rate will tend to be negative for short lags, and the estimates for the spectrum will tend to rise as frequency increases, indicating that the exchange fund has increased the variance in the percentage changes by increasing the day-to-day and week-to-week movements relative to the year-to-year movements in the rate.

II. *Evidence*

The period of flexible rates from January, 1952, to November, 1960, contains approximately 2,200 trading days.[9] Table 2 gives the autocor-

TABLE 2

AUTOCORRELATION ESTIMATES FOR PERCENTAGE CHANGES
IN DAILY RATE, ONE TO FIFTY DAY LAG
JANUARY, 1952, TO NOVEMBER, 1960

Lag	Estimate	Lag	Estimate
1	0.200*	26	−0.016
2	0.035*	27	−0.024
3	0.016	28	−0.008
4	0.065*	29	0.010
5	0.043*	30	0.016
6	0.002	31	0.003
7	0.027	32	−0.008
8	0.039*	33	0.012
9	0.035*	34	0.035*
10	0.057*	35	0.005
11	−0.009	36	−0.043*
12	−0.011	37	−0.026
13	0.026	38	−0.004
14	0.047*	39	0.003
15	0.022	40	−0.025
16	−0.004	41	−0.018
17	0.015	42	0.007
18	−0.019	43	−0.015
19	−0.005	44	−0.017
20	0.048*	45	−0.021
21	0.020	46	0.011
22	0.019	47	0.017
23	−0.014	48	−0.000
24	0.001	49	0.011
25	0.024	50	0.011

* Significant at the 90 percent level.

[9] The data used in this analysis are the U.S. prices of Canadian dollars as given in the *Wall Street Journal*. The amount of data precludes including them as part of the article. However, the data will be supplied to any bona fide academic research project for the cost of the cards necessary to duplicate the data deck.

relation estimates up to a fifty day lag for the percentage changes in
the daily rate during the nine years. The pattern is clear. The esti-
mates for short lags are positive and for four out of the first five lags
the estimates are significant at the 90 percent level. This indicates a
strong tendency for sustained movements in the rate and supports the
claim that there was a strong stabilizing influence in the exchange mar-
ket.

Figure 3 shows the estimates of the spectrum for the percentage
changes in the daily rate from January, 1952, to November, 1960,
plotted against frequency. Once again the pattern is clear. The higher
the frequency, i.e., the shorter the cycle, the smaller is the estimate for
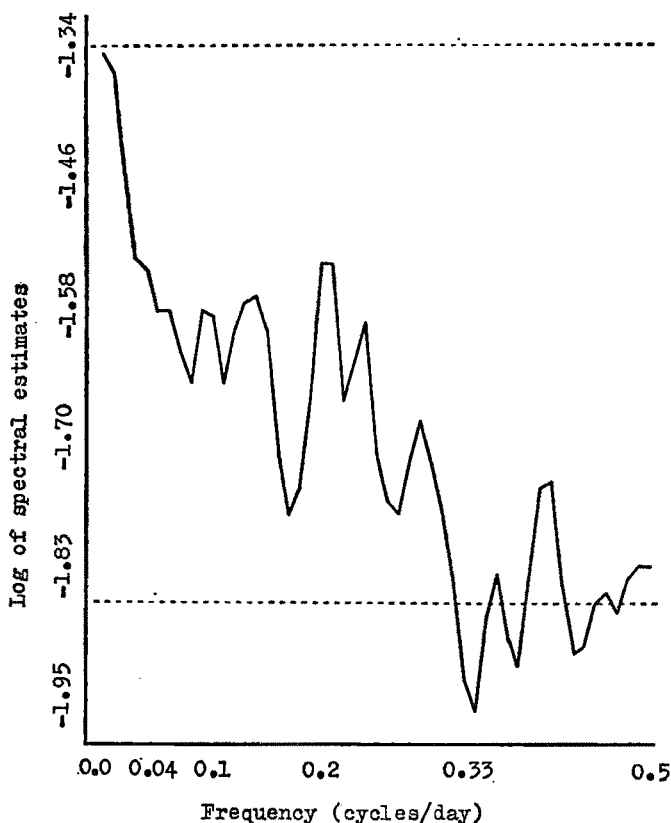the spectrum. This indicates that the day-to-day and week-to-week

FIGURE 3

SPECTRAL ESTIMATES PLOTTED IN LOG SCALE
AGAINST FREQUENCY, PERCENTAGE CHANGES IN
DAILY U.S. PRICE OF CANADIAN DOLLAR
(JANUARY, 1952, TO NOVEMBER, 1960)

movements in the rate contributed relatively less to the variance in the percentage changes in the rate than the month-to-month and year-to-year movements in the rate. As in the case of the autocorrelation estimates, this is the pattern expected from an influence in the market that reduces the short-run movements in the rate without materially affecting long-run movements.

III. *Conclusions*

Given the results of the autocorrelation and spectral analysis, there can be little doubt that something was stabilizing the Canadian exchange rate during the period of flexible rates. Was this stabilizing influence speculation or the Canadian Exchange Fund? There are at least two reasons for choosing the Exchange Fund. First, the Exchange Fund was publicly committed to ironing out short-run fluctuations without preventing long-run movements in the rate. Second, European countries which had freely fluctuating rates without severe inflation during the early 1920's, e.g., England, France, and Spain, show only a slight tendency for autocorrelation estimates for percentage changes in daily rates to be positive for short lags and no noticeable tendency for spectral estimates of percentage changes either to increase or decrease as frequency increases.[10] However, the Norwegian rate during the early 1920's, which was flexible but was influenced by central bank activity designed to eliminate "unsound speculation," shows a pattern for the autocorrelation function and spectrum of the percentage changes in the daily rate that is almost identical to that shown above for the Canadian experience.[11]

The evidence strongly suggests that the Canadian Exchange Fund was successful in its attempt to iron out short-run fluctuations in the rate. Neither spectral nor autocorrelation analysis can determine whether or not the Exchange Fund reduced long-run movements as well as short-run movements. These tools can indicate only that short-run fluctuations were reduced relative to long-run movements. However, the small rise in official Canadian holdings of gold and U.S. dollars from January, 1952, to November, 1960 ($54 million or 3 percent) indicates that the Exchange Fund did not attempt to resist long-run movements in the rate.[12]

The fact that an exchange fund can reduce short-run fluctuations

---

[10] Pippenger, *op. cit.*, pp. 314-19.

[11] "There has been no interference, either by legislation or by administrative measures, with dealings in foreign exchange and the market has accordingly been left free. The Norges Bank, Bank of Norway, has, however, felt obliged to play an active part in the market in order to prevent unsound speculation in exchange." N. Rygg, Managing Director of the Bank of Norway, "The Currency Problem in Norway," in John P. Young, *European Currency and Finance* (Washington, D.C.: G.P.O., 1925), Vol. I, p. 216.

[12] International Financial Statistics, International Monetary Fund.

without materially affecting long-run movements does not imply, by itself, that any future experiment with a flexible rate should include a stabilization fund. Other things being equal, a stable rate is preferred to one that fluctuates more, but stability is not the only criterion for determining the efficiency of a foreign exchange market as a price generating mechanism. A market that never permitted the price to move would be a poor pricing mechanism indeed.

DISCUSSION

HARRY EASTMAN: When economists get around to measuring the total welfare implications of the Canadian experience with flexible rates of exchange, they must not omit on the positive side the source of data for research that that experience provided and the contributions to knowledge that research has made. Professor Pippenger's work is an instance of this, for he has added very substantially to our understanding of the effects of operations to stabilize exchange rates under certain conditions. He made certain assumptions about the character of the foreign exchange market and then shows that an exchange stabilizing authority that leans against the wind by selling when the price of foreign exchange rises and buying when it falls reduces short-term fluctuations in the rate relative to long-term fluctuations while attenuating both. Exchange stabilization funds do in fact usually operate according to the rules envisaged in this model. The reduction of short-term fluctuations relative to long-term ones stems from a sort of averaging process which itself stems from the independence of each day's private demand and supply conditions from the volume and price of the previous day's transactions. The SD curve hops up and down randomly and gives rise to random price movements in the absence of official intervention. If a price change is reduced by intervention, that leaves over a bigger or smaller potential price change for the next day which may in turn be reduced. In this context, intervention introduces serial correlation, and, as the previous day weighs most in each day's average, the autocorrelation is highest for adjoining days so that short-term fluctuations are reduced relative to long-term ones.

The assumption of independence of a day's SD curve from whatever may have happened the previous day is perfectly legitimate, but we must note that an equally simple assumption could be made that would give different results. This point is made lest we go out and prematurely attempt to stabilize all prices according to the simple rule of resisting all changes in those prices. If random changes in the price took place from one day to the next but the change departed from the price as it existed the previous day, intervention that reduced each change would not introduce serial correlation and would not therefore reduce short-term fluctuations relative to long-term fluctuations. If the series r_t, $r_t + e_{(t+1)}$, $r_t + e_{(t+1)} + e_{(t+2)}$, etc., are not correlated, neither is the series

$$r_t, r_t + \frac{e_{(t+1)}}{2}, \qquad r_t + \frac{e_{(t+1)}}{2} + \frac{e_{(t+2)}}{2} \text{ etc.}$$

In this case, the spectrum of the series for the stabilized rate as for the unstabilized rate would appear as a horizontal line and a comparison with the spectrum of the actual rates would not reveal stabilizing activity even if it existed.

My cautious gaze now turns to the spectrum of the series of actual rates of exchange and to wonder if it is entirely plausible that the long-term fluctua-

tions that lasted many days, indeed months, were created by random forces and not by systematic forces, such as seasonality, and by nonstationary factors. The market for foreign exchange is highly competitive it is true, but the period analyzed is a long one during which major changes in conditions occurred and these could not have been foreseen. The first half of the period was generally prosperous, the second depressed. The capital inflow that began shortly after the beginning of the period was mostly destined to finance new private investment at the beginning whereas later it was composed much more largely of provincial and municipal borrowings. It seems to me that the most reasonable view to take is that the long fluctuations actually taking place in the rate were not the result of accident. They were after all both long and very large relative to day-to-day rate changes. This consideration cuts pretty deep, of course, because if the long-term fluctuations are not the result of random changes, but reflect systematic changes, their large size relative to the short-term fluctuations, which likely were the result of random forces, should not be explained by the possible intervention of the Exchange Fund Account.

I should like to emphasis that my remarks are in no way disproof of Professor Pippenger's thesis but are an attempt to put his extremely interesting results into a somewhat wider context than he was able to do in a brief paper. They do perhaps introduce some caution in estimating the role of the Exchange Fund Account, a caution reinforced when we take into consideration the possibility of a trend in the rate.

The consideration of a possibility of changes in the long-term trend in the rate of exchange can best be done by means of the SD function. The first case is that in which SD moves to $S'D'$ and stays there, the long-term equilibrium rate of exchange having changed from R_1 to R_2'. Intervention slows the movement to R_2'. Such intervention is called a "stabilizing influence" in the paper even though the amplitude of the fluctuation is not reduced. This view is tantamount to saying that the reduction in the rate of the adjustment process to a real change in conditions is desirable, with which I do not agree.

$S'D'$ may not stay up there eternally, of course. It may begin to fall again. In that case, the fluctuation will be reduced, the reduction being less the more days pass before $S'D'$ moves downward. But these are different models from Pippenger's statistical model, because in these models serial correlation exists between the position of $S'D'$ on the consecutive days both in the absence and in the presence of stabilizing activity.

Consider the consequence of yet another case: that in which $S'D'$ rises continuously. Leaning against the wind leads the stabilization fund to fall flat on its face, if I may elaborate the metaphor. It either accumulates or decumulates reserves to some limit of practicability. And that is why funds intervene relatively gingerly in foreign exchange markets when they are genuinely trying to smooth out short-term fluctuations in the rate rather than influencing its level.

The Canadian Exchange Fund Account appears to have learned of the effects of changes in trend in the rate on reserves in 1951-52 when the rate altered from fluctuating about a price of roughly $1.05 for a U.S. dollar to around $.97. During the first five months of the slide in the rate, the Exchange

Fund Account seems to have intervened at a rate of about $130 million for each cent change and in the second five months at a rate of only $7.5 million for a change of one cent. These latter months form the beginning of Professor Pippenger's period in which it turned out, but this was not forseeable, that there was a zero trend in the rate over the entire period. It is very difficult to deduce from public data what was the extent of Exchange Fund Account intervention with movements in the rate, but the inconsiderable month-end to month-end changes plus the evidence of reluctance to interfere with trend may well mean that the assumption of halving a variances in the rate overstates any possible actual performance. This latter set of considerations seems to me to offer a field for investigation the importance of which is best seen in the context of Professor Pippenger's major contribution to our knowledge of the period.

DALE W. JORGENSON: One of the most fruitful sources of new theoretical ideas in economics has been the reconsideration of atemporal equilibrium problems from the intertemporal point of view. Sidrauski's paper is an interesting contribution to this enterprise. His point of departure is the standard theory of money as developed especially by Patinkin. In this theory money is treated as a durable consumers' good entering directly into the utility function in the form of money services.

In Patinkin's version of the theory the problem is to determine equilibrium in the markets for money, bonds, goods, and labor. Sidrauski proposes to eliminate bonds from consideration and to determine equilibrium time paths for accumulation of stocks of money and goods and for consumption of goods and money services. The fundamental tool is the maximization of an intertemporal utility function for a representative individual, subject to a fixed technology for production and a predetermined government policy for creation of money.

The step from atemporal to intertemporal theory is a short one for Sidrauski, thanks to the special character of technology and the intertemporal utility function. Utility is simply a weighted sum (or integral) of instantaneous utilities while technology depends only on current inputs, including the input of capital services, where capital is a weighted sum of past investments. Accordingly, the single firm in the economy equates marginal productivities of capital and labor to corresponding real prices of these services; the household equates marginal utility of goods to the price of these goods and the marginal utility of money services to their price.

These parts of Sidrauski's construction correspond directly to atemporal theory. The model is closed by adding two conditions: (1) the subjective rate of discount is equal to the money rate of interest; (2) as the stock of assets grows over time, the limit of the discounted nominal value of this stock per capita must be zero.

At this point two relationships with atemporal theory may be established. First, the results of atemporal theory may be interpreted as a characterization of an instantaneous equilibrium along a dynamic time path. Second, atemporal theory may be interpreted as a theory of the stationary state; that is, an instantaneous equilibrium that persists through time. Despite the nearly complete correspondence between the intertemporal theory outlined above and the

standard atemporal theory, neither of these relationships is developed with any precision by Sidrauski.

To give some indication of problems remaining to be clarified, consider Sidrauski's equation (20), where quantity demand of real capital is "a function of the implicit interest rate." This function must be studied in the context of identities (7) and (8), defining investment as the rate of change of capital plus replacement. If the interest rate changes, capital changes by a discrete amount, which means that investment and, hence, society's output must be unbounded. Thus, such a function may only be employed to characterize alternative stationary states while Sidrauski seems to suggest that this relationship pertains to movement along a dynamic equilibrium path.

Another problem of this type is the interpretation of the initial conditions for differential equations in stocks of money and goods. Sidrauski writes that the time path is to be found, "given the initial stock of real wealth." But what determines the division of this stock between money and goods? Is this to be taken as given at the outset as part of the initial conditions or will some *deus ex machina* provide society with opportunities to buy goods and sell money (or vice versa) in the amounts required to put the social portfolio into initial balance?

Finally, in Sidrauski's model we have at least three prices for goods: the price p_t, the implicit price λ_t, and the expected price. Are all of these prices the same? If not, why can society predict according to the first-order differential equation (31) when it is unable to predict according to the first-order differential equation (16), which is always perfectly accurate? One of these equations, the equality between the money rate of interest and the social rate of discount, appears to be a differential equation that makes sense only if the initial price is a certain function of the initial stock of assets. What if the price is something else?

This is only a partial list of questions which might be raised about the economic implications of Sidrauski's extensive computations. From a more formal point of view, other questions might be suggested, such as, what are sufficient conditions for solution of the problem of intertemporal utility maximization? The conditions set down are necessary if nonnegativity of the asset stocks is ignored; in a more thorough analysis, these conditions are not even necessary.

Turning briefly to the conclusions of the paper, we read that "the higher the rate of monetary expansion the lower will be the steady state level of utility." Of course this statement implies that the lower is the rate of monetary expansion the higher will be the steady state level of utility so that monetary and fiscal policy should be devoted to taxing consumers and salting away the cash proceeds in society's bank vaults, the faster the better.

We conclude that Sidrauski has opened up an interesting new area of research and that much remains to be done, both in setting out a useful framework for intertemporal analysis and in establishing the relationship between intertemporal theory and established results of atemporal theory.

In pursuing these objectives perhaps theorists should not be too traditional in their approach to formal analysis. Simply writing down a few conditions for equilibrium and counting equations and unknown may have been sufficient for

many atemporal problems, but atemporal theorists can draw on an economic interpretation of the corresponding formalism that has been carefully worked out and refined over the nearly 130 years since Cournot. In intertemporal theory greater care and deeper analysis are mandatory, at least until the economic interpretation of a carefully executed formal analysis is made completely explicit.

PETER O. STEINER: As will become clear at once, I am somewhat critical of Rosse's paper. It is probable that my comments would be much changed if I had seen the dissertation itself, since I am aware of how difficult it is to do justice to a major empirical study in a short paper. The substance of Rosse's work concerns the estimation of marginal costs in an industry with a complex output. The paper (as distinct perhaps from the dissertation) strives to place this work in a monopolistic competition framework and to explain thereby the disappearance of competition among newspapers.

An unkind paraphrase of the paper might read: "I have rather cleverly fitted a model (which I needn't bother to specify) that makes it difficult to reject the hypotheses of declining marginal cost. Since declining marginal cost implies declining average cost, I infer economies of scale. Economies of scale are sufficient to explain isolated one-newspaper markets."

In fact, Rosse says that economies of scale are necessary, not that they are sufficient, to the existence of one-newspaper towns. If he genuinely believed the necessity of scale economies, he could have saved himself all that work: the well-documented presence of so many one-newspaper towns would establish the necessary existence of economies of scale in newspaper production. His whole paper clearly portrays the belief that his cost findings carry a long way toward explaining the phenomenon of the one-newspaper town.

I have serious reservations concerning declining marginal costs as the explanation of the one-newspaper town. On the one hand, why should they not lead to one big regional newspaper supplemented by "$\frac{1}{7}$ newspaper towns" (i.e., weeklies), and other news media? Is it demands of advertisers or readers which make the optimal number 1 rather than 2, or zero, or $\frac{1}{7}$? On the other hand, we are left in doubt as to why there were so many multicity newspapers in 1923? Were there no economies of scale then? If there are definable changes that have occurred, perhaps one should define them and relate the time series of newspapers per city to the hypothesized causal factors. Are such scale economies the result of technological changes, or of the rise in labor costs, or of restrictive practices of labor unions—or of all three? In any case analysis of changes over time would add greatly to our understanding of Rosse's data.

Put a bit more formally, I would expect that the number of independent newspapers in a community would be an argument (with positive partial derivative) in the demand for newspaper output as well as (with negative partial derivative) in cost. It may well be that outside changes have transformed these functions so that there is a kink at $n = 1$. But declining marginal costs are neither necessary nor sufficient to this result.

If a Chamberlinian model is used here, I fail to find it. Any situation char-

acterised by product differentiation may, I suppose, be classified as monopolistically competitive; but does it help to do so? As a critical test I find here no comfort either to the true believers or to the critics of that theory.

These reservations about the framework in which Rosse presents his empirical results do not detract from the results themselves. They appear to reflect both ingenuity and competence. Because Rosse has not given us his basic model it is impossible to form a reliable judgment. One looks forward to a more detailed write-up of this promising piece of research.

INVITED DOCTORAL DISSERTATIONS II

NATURAL RESOURCES, FACTOR MIX, AND FACTOR REVERSAL IN INTERNATIONAL TRADE*

By SEIJI NAYA

University of Wisconsin

Most empirical tests of pure trade models are for those employing two factors, capital and labor. An example is the celebrated work of Leontief on the structure of U.S. foreign trade, the results of which are termed "The Leontief Paradox."[1] Many writers have been critical of Leontief's omission of a third group of factors, natural resources.[2] The possibility that the paradox is largely due to natural resources (R) is given by Leontief in one of his computations which excludes 19 R sectors (in his second article). Vanek's study on the role of R in U.S. trade provides additional empirical support to the R-explanation of the paradox.[3]

Vanek's study, however, is based only on U.S. data. This paper tests the applicability of his explanation of the paradox to the trade structure of Japan, India, and Canada (Section I). The capital (C), labor (L) and R product requirements of Japan's foreign trade are computed, but for India and Canada, the findings of Bharadwaj and Wahl are used.[4] The conclusion of the first section is then used to evaluate Minhas' ranking test on C-L intensity reversals, another explanation given for the paradox (Section II).[5]

[1] "The Leontief Paradox" refers to the findings of W. W. Leontief that exports of the U.S. contain relatively less capital but more labor than domestic replacements of its competitive imports (in "Domestic Production and Foreign Trade: The American Capital Position Re-examined," *Proceedings of the Amer. Philos. Soc.*, Sept., 1953, pp. 331-49; and "Factor Proportions and the Structure of American Trade: Further Theoretical and Empirical Analysis," *Rev. of Econ. and Statis.*, Nov., 1956, pp. 386-407).

[2] B. C. Swerling, "Capital Shortage and Labor Surplus in the United States?" *Rev. of Econ. and Statis.*, Aug., 1954, p. 289; I. B. Kravis, "Availability and Other Influences on the Commodity Composition of Trade," *J.P.E.*, Apr., 1956, pp. 143-55; G. 'Haberler, *A Survey of International Trade Theory* (Int. Fin. Sec., Princeton Univ., 1961), pp. 21-23.

[3] Vanek, *The Natural Resource Content of United States Foreign Trade, 1870-1955* (Cambridge, 1963), especially pp. 132-35.

[4] R. Bharadwaj, "Factor Proportions and the Structure of Indo-U.S. Trade," *Indian Econ. J.*, Oct., 1962, pp. 105-16, and *Structural Basis of India's Foreign Trade*, Series in Monetary and International Econ., No. 6, Univ. of Bombay, 1962; Donald F. Wahl, "Capital and 'Labor Requirements for Canada's Foreign Trade," *Canadian J. of Econ. and Polit. Sci.*, Aug., 1961, pp. 349-58.

[5] B. S. Minhas, *An International Comparison of Factor Costs and Factor Use* (Amsterdam, 1963), pp. 39-42.

I

In measuring the R factor, a proxy method is employed by Vanek—the R products embodied in a unit value of a given commodity are substituted for the R factor requirement under the assumption that the former is proportional to the latter. Twenty-one sectors (various agricultural and mining sectors and the logging sector) out of 192 intermediate sectors are selected and their outputs are used to determine the role of the R factor in production and trade.

The natural resource product requirements per million dollars of American exports (R^x) and competitive import replacements (R^m) computed by Vanek are defined as follows[6] and given in Table 1 along with Leontief's C and L requirements: $R^x = BX$ and $R^m = BM$ where

TABLE 1

DOMESTIC CAPITAL, LABOR, AND NATURAL RESOURCE PRODUCT REQUIREMENTS
PER MILLION DOLLARS OF AMERICAN EXPORTS AND OF
COMPETITIVE IMPORT REPLACEMENTS, 1947

	Exports (X)	Imports (M)	X/M
1. Capital (dollars)...................	2,550,780	3,091,339	0.83
2. Labor (man-years)................	182,313	170,004	1.07
3. Natural resource products (dollars)...	340,000	630,000	0.54

SOURCE: Vanek, *op. cit.*, p. 132.

B refers to the 21 by 192 portion of the inverse matrix while X and M refer to the proportionate distribution of a million dollars, each, of U.S. exports and competitive imports, respectively.

As can be seen from the table, import replacements contain a considerably larger R output than do exports. Since R products are assumed, on the average, to be more R factor-intensive than nonresource products, this difference in R^m and R^x implies that the U.S. engages in trade to economize her scarce factor, R.

Vanek's finding on R^m and R^x does not directly explain the paradox. To do this, he relies on the hypothesis that R and C are a pair of complementary factors while R and L are not; that is, the use of R would jointly require a relatively large amount of C. Hence, C may be the relatively more abundant factor compared to L. But being complementary with R, more of it is also needed to replace U.S. imports than for exports.

Although Vanek's explanation of the paradox is based on the critical "complementarity" between R and C, the only test provided for this proposition is his general observation that the R product content and

[a] Vanek, *op. cit.*, pp. 82-83.

capital requirements of commodities are positively related while the R product content and L requirements are not.[7] His finding then implies that the C-L ratio of R products tends to be relatively high in comparison with nonresource products as is generally shown in Leontief's C-L figures of U.S. industries.

One way to test the generality of Vanek's analysis of the paradox is to examine its applicability to the trade structure of other countries. If C and R are complementary in the U.S., they must also be complementary in other countries, assuming identical production functions. The C-L structure of foreign trade of a country such as Japan, which is considered to be even more characterized by scarcity of R than the U.S., should, under this assumption, conform to the Leontief paradox while the reverse should hold true for a country such as Canada.

In order to compare the factor structure of U.S. foreign trade with that of Japan, the natural resource product requirement (R^x and R^m) for 1955 and 1959 (Table 2) is computed according to Vanek's ap-

TABLE 2

THE NATURAL RESOURCE PRODUCT REQUIREMENT PER MILLION YEN OF JAPANESE EXPORTS AND COMPETITIVE IMPORT REPLACEMENTS, 1955 AND 1959

Year	Exports (yen)	Imports (yen)	Exports / Imports
1955	238,990	910,043	0.263
1959	221,051	742,085	0.298

proach. The inverse matrix (as well as direct capital and labor coefficients used later) is based on 35 intermediate sectors.[8] In computing the R^x and R^m, the four sectors treated as R sectors are: Agriculture, Forestry, and Fishery; Coal and Lignite; Crude Petroleum and Natural Gas; and Metals and Nonmetallic Minerals.

The results of Table 2 clearly reflect the generally accepted scarcity of R in Japan's economy. The ratio of R^x and R^m is 0.263 (1955) and 0.298 (1959). Although the American and Japanese interindustry classifications are not strictly comparable, the figures for Japan are still strikingly lower than the 0.54 for the U.S.

If R and C, but not R and L, are complementary, then the relative requirements of C for Japan's import replacements should be larger than for her exports. Even without the complementarity proposition, such a result for Japan would not be surprising since her comparative

[7] *Ibid.*, p. 133.
[8] These data are given in *Interindustry Analysis of the Japanese Economy* (based on the 1955 interindustry table), Japan, Ministry of International Trade and Industry, Minister's Secretariat, Div. of Res. and Statis., 1962.

advantage is generally assumed to lie in relatively L-intensive products. But the computations on the C-L requirements of Japanese foreign trade for 1955 and 1959 (which are in line with Leontief's) do not support the above contentions (see Table 3, Row A). For both years tested, Japanese exports require relatively more C and less L in comparison with import replacements.[9]

TABLE 3

RESULTS OF CAPITAL AND LABOR REQUIREMENTS PER MILLION YEN OF JAPANESE
EXPORTS AND COMPETITIVE IMPORT REPLACEMENTS, 1955 AND 1959

Coverage of Sectors	Years		Capital ¥ 1,000	Labor (man-years)	C/L	Export Ratio —————— Import Ratio
A. All sectors..........	1955	Competitive Imports	1,206	5.119	236	1.678
		.Exports	1,289	3.258	395	
	1959	Competitive Imports	1,265	3.164	400	1.315
		Exports	1,504	2.859	526	
B. Excluding natural resource sectors.....	1955	Competitive Imports	703	1.352	520	0.985
		Exports	1,125	2.197	512	

R being scarce for Japan, if R and C are complementary, then the exclusion of the R sectors should increase the C-L ratio for Japanese exports relative to that for import replacements. For, if R industries were relatively capital intensive, the exclusion of these industries would tend to exert a larger downward influence on the C-L ratio of import replacements than on exports. But the result of the computation excluding the R sectors for 1955 (Table 3, Row B) shows that the C-L ratio of both import replacements and exports has gone up. More significantly, the C-L ratio of the import replacements has risen sufficiently more than that of exports to make the former ratio slightly greater than the latter.

It is interesting to note that for both the U.S. and Japan, which are assumed to be scarce in R, the paradoxical findings on the C-L structure of trade are reversed and the expected results from the Heckscher-Ohlin hypothesis are derived when R sectors are excluded in the C-L

[9] Similar findings for 1951 are shown in M. Tatemoto and S. Ichimura, *Factor Proportions and Foreign Trade: The Case of Japan* (Discussion Paper No. 8, Inst. of Soc. and Econ. Res., Osaka Univ., 1958). Reprinted in *Rev. of Econ. and Statis.*, Nov., 1959. pp. 442–46.

computation. However, there is a sharp difference in the direction of
this change in the *C-L* requirements of the two countries. The *C-L*
ratio of exports has gone up in comparison to import replacements for
the U.S. by excluding *R* sectors (Leontief's study), while the reverse
is true for Japan. If *R* and *C* are complementary factors everywhere,
exclusion of *R* sectors should have reduced the complementary factor,
C, more from import replacements than exports for both countries. Al-
though the U.S. data appear to support the complementarity proposi-
tion, the Japanese results are clearly inconsistent with it.

The *C-L* ranking of Japanese industries (Table 4) shows that three
out of four *R* sectors have relatively low *C-L* ratios. The agricultural
sector has the lowest ratio of all industries both in 1955 and 1959. In

TABLE 4

CAPITAL-LABOR RATIOS FOR JAPANESE INDUSTRIES AND THEIR RANKINGS, 1955 AND 1959

Industry	Ranking*	
	1955	1959
1. Agriculture, forestry, and fishery	35	35
2. Coal and lignite	25	28
3. Crude petroleum and natural gas	6	7
4. Metals and nonmetallic minerals	18	23
5. Processed food	30	30
6. Natural textile mill products	26	26
7. Synthetic textile mill products	9	6
8. Textile and fabric products	27	24
9. Paper and pulp	22	18
10. Lumber and wood products	33	34
11. Miscellaneous manufacturing	28	25
12. Coal products	14	21
13. Petroleum products	3	3
14. Chemical fertilizers	11	11
15. Basic and intermediate chemical products	8	9
16. Final chemical products	19	15
17. Nonmetallic mineral products	16	16
18. Primary iron and steel products	4	4
19. Steel casting, forging, and rolling	5	5
20. Finished metal products	13	27
21. Nonferrous metals	7	8
22. General industrial machinery	23	22
23. Precision machinery	29	33
24. Industrial electrical machinery	17	20
25. Nonindustrial light electrical machinery	21	17
26. Automobiles	15	14
27. Ships and locomotives	20	19
28. Construction	31	31
29. Engineering	34	32
30. Trade	32	29
31. Transportation	2	2
32. Electricity	1	1
33. Other services	12	12
34. Scraps	10	10
35. Unclassified	24	13

* Rankings based on total factor requirements.

contrast, crude petroleum is relatively capital intensive, 6th and 7th in the C-L ranking in 1955 and 1959, respectively. This sharp difference in the ranking of agriculture and crude petroleum does not by itself justify complete rejection of the complementarity proposition, but it certainly provides additional evidence that Vanek's observations on the relationship between R and C cannot be generalized without qualification. The imports of the four resource sectors are very large while the exports are extremely small. But, among the sectors, the net imports of agriculture are by far the largest and hence the result that Japan exports relatively C-intensive products and imports L-intensive goods (Table 3, Row A) must have been strongly influenced by the large net imports of the agricultural sector and its low C-L ratio.

Two separate studies of India and Canada add to the evidence on the influence of R on the C-L structure of trade. Bharadwaj's study of India shows that the composite C-L ratio of Indian total exports is less than that of the replacements of its total competitive imports.[10] However, when the C-L requirements of the country's bilateral trade with the U.S. are considered, the results are reversed: Indian exports to the U.S. are shown to be relatively C-intensive and imports from the U.S., L-intensive.[11]

Wahl's study on the C-L structure of Canadian trade also gives an apparently contradictory result: C-intensive exports but L-intensive imports.[12] His finding may not necessarily be inconsistent with the Heckscher-Ohlin hypothesis since Canada may well be relatively C-abundant vis-à-vis the rest of its trading partners. But even in Canada's bilateral trade position with the U.S. and England, he stills gets a result similar to that for total trade.

Each writer appropriately points out that the relatively large share of primary products in the commodity composition of his country's trade might have influenced the paradoxical findings on the bilateral trade positions. Bharadwaj reasons that the normally large Indian imports of agricultural products from the U.S., which were even larger in 1951 (the year tested) due to a crop failure, tended to lower the C-L ratio of imports from the U.S. while relatively large exports of minerals and ores tended to raise the C-L ratio of Indian exports to the U.S. Wahl similarly qualifies his result in terms of a large share of primary products in Canadian exports and a predominance of manufactured goods in its imports. Raw material-oriented exports of Canada, which are C-intensive, tend to raise the C-L structure of exports but imports, oriented toward manufactured goods which are L-intensive, tend to lower the C-L ratio of competitive imports.

[10] Bharadwaj, *Structural Basis of India's Foreign Trade*, pp. 56-57.
[11] Bharadwaj, "Factor Proportions and the Structure of Indo-U.S. Trade," p. 109.
[12] Wahl, *op. cit.*, p. 353.

The somewhat paradoxical results shown for the U.S., Japan, India, and Canada all point to one common feature: the significant influence of R products in the C-L structure of each of their trade patterns. More importantly, however, the influence seems to emerge in two distinctive patterns associated with different kinds of R products originating from agricultural and extractive industries. The influence of extractive industries is somewhat in line with the complementarity proposition while that of agricultural industries directly contradicts the proposition in explaining the paradoxical results of the four countries considered.

What are the implications of the above analysis? For a country such as Canada, the C-L requirements of its composite exports tend to be C-intensive since exports are raw material-oriented and her import requirements L-intensive because of large imports of manufactured goods. On the other hand, for a country like Japan, exports are also likely to be C-intensive not simply because the country is capital rich but because it is land scarce and imports large amounts of agricultural products which are extremely L-intensive (in contrast to the C-intensity in the U.S. and Canada). Similar reasoning can be applied to explain both the Leontief paradox and the Indian bilateral trade with the U.S.

Vanek's complementarity explanation may apply to the U.S. and Canada as both agricultural and extractive natural resources are C-intensive in these countries, but it does not apply to countries such as Japan and India because of the contrasting C-L structure of their agricultural sectors. The specific natural resource required in agricultural processes, i.e., land, may very well not be complementary with capital.

II

The main purpose of the last section was to point out that consideration of capital and labor alone is inadequate to explain trade patterns since the exchange of R products may be dictated more by the R factor. Such reasoning can also be directed to the testing of the C-L intensity reversal when the R sectors are among the industries compared, as in one of the two tests employed by Minhas.

Because of the small number of industries tested in employing the CES production function.[13] Minhas extends his work to include a relative ranking comparison of the C-L ratio of 20 industries to the U.S.

---

[13] In demonstrating factor reversal or crossover points, Minhas selects only 6 of the 24 industries and finds 5 out of 15 possible crossover points. In evaluating this work, Leontief made computations for 21 of the same industries, basically using the data furnished in Minhas' work, in "International Factor Cost and Factor Use," *A.E.R.*, June, 1964, Part I, pp. 54, 335-45. Interestingly, Leontief's result hardly supports the generality of factor reversal claimed by Minhas. Out of 210 possible crossover points, only 17 occur in Leontief's computation.

(1947) and Japan (1951). The *C-L* ranking for these 20 industries are reproduced in Table 5.

Minhas compares the relative ranking of *C* intensities based on total (direct and indirect) factor requirements and also on just the direct requirements. The Spearman rank coefficient of correlation between the two orderings (U.S. and Japan) based on total factor requirements is only .0328.[14] Such a low value definitely implies a large dissimilarity of relative *C* intensities, supporting the factor reversal argument. However, for the direct factor requirements only, Minhas finds a greater similarity (coefficient value of 0.73).

Before identifying the source of this low coefficient, a few qualifications should be raised. First, the industries compared are too broadly aggregated to provide a conclusive result for the purpose of inter-country comparison. It seems there exists a large difference in the classification of industries. For example, the food processing sector in the Japanese input-output table includes tobacco and cigarette manufacturing while it does not in the U.S. input-output table. Questions of comparability also arise in the case of apparel and machinery. In view of these likely differences in the commodity mix for the industries treated as identical, it seems more realistic to evaluate the overall similarity or dissimilarity than a one-to-one correspondence of the factor intensity ranking.

Even if one refrains from comparison of a one-to-one correspondence, the orderings based on the total factor intensity cannot be regarded as similar. It is clear from Table 5 and from Section I that the largest difference in the *C* intensity is found in agriculture. While the total *C* intensity is ranked 3rd in the U.S., it is ranked 20th (lowest) in Japan, or a difference of 17. Therefore, agriculture exerts the most dominant downward influence on the value of the Spearman coefficient.

Notice further that the other land intensive sectors, e.g., grain mill and processed food, are also far apart in the *C-L* ranking in terms of total factor requirements (a difference of 15 and 8, respectively). But such large differences are not shown in the *C-L* ranking of the direct requirements for these two sectors (a difference of only 3 for each). This is to be expected, however, since these two sectors tend to require relatively large inputs of agricultural products in their production processes. On the one hand, the use of *L*-intensive agricultural inputs in Japan for the production of grain mill products and processed food would tend to lower their total *C*-intensity ranking in comparison with the ranking of their direct factor requirements. On the other hand, the *C*-intensive agricultural inputs required for these sectors in the U.S. raises the total factor ranking in comparison to the direct factor rank-

[14] Minhas, *op. cit.*, pp. 39-40.

TABLE 5

RANKING OF INDUSTRIES BY CAPITAL INTENSITY FOR THE U.S. (1947) AND JAPAN (1951)

Industry	Ranks Based on Total Capital and Labor Requirements			Ranks Based on Direct Capital and Labor Requirements		
	U.S.	Japan	Difference	U.S.	Japan	Difference
Petroleum products.........	1	1	0	1	1	0
Coal products..............	2	2	0	2	2	0
Agriculture................	3	20	−17	3	14	−11
Grain mill products.........	4	19	−15	9	6	3
Processed foods............	5	13	− 8	10	7	3
Chemicals.................	6	5	1	6	4	2
Nonferrous metals..........	7	4	3	4	3	1
Iron and steel..............	8	3	5	5	5	0
Paper and products.........	9	11	− 2	7	15	− 8
Nonmetallic mineral products.................	10	9	1	8	11	− 3
Textiles...................	11	15	− 4	18	12	6
Transport equipment.......	12	10	2	11	9	2
Machinery.................	13	6	7	12	10	2
Rubber and products.......	14	12	2	14	16	− 2
Ship building..............	15	7	8	13	8	5
Lumber and wood..........	16	17	− 1	15	17	− 2
Industry not elsewhere classified................	17	16	1	17	20	− 3
Printing, publishing.........	18	8	10	16	18	− 2
Leather...................	19	18	1	19	19	0
Apparel...................	20	14	6	20	13	7

SOURCE: Minhas, *op. cit.*, Table V, p. 40.

ing. As shown in the table, the rankings for grain mill products and processed food in Japan are 6th and 7th, respectively, based on the direct requirements but are lowered to 19th and 13th based on the total requirements, while they are raised in the U.S. from 9th and 10th to 4th and 5th in rank. Therefore, the wide difference in the orderings based on the total factor requirements obtained by Minhas, implying the strong possibility of factor reversal, is largely due to the direct influence of agriculture and the indirect influence of agriculture on grain mill products and processed food.

In order to minimize the likely bias from the land factor, a *C-L* ranking comparison excluding these three sectors has been made.[15] The result indicates a significant improvement in the similarity of the rankings. The value of the Spearman coefficient is now 0.826 for the direct *C-L* requirements and 0.765 for the total factor requirements (compared to 0.73 and 0.328, respectively, when the three sectors are included). The rankings for both direct and total factor requirements still do not match exactly, but the degree of similarity between them

[15] A recent article also evaluates Minhas' ranking test by exclusion of such sectors (D. S. Ball, "Factor-Intensity Reversals in International Camparison of Factor Costs and Factor Use," *J.P.E.*, Feb., 1966, pp. 77-80).

is sufficient to make one question the factor reversal argument on the basis of this ranking test.

III

Emphasis in this paper on the influence that natural resources exert on the C-L structure of trade leads to a final conclusion that empirical applications of the factor proportions theory will be more meaningful when factors in addition to capital and labor are explicitly incorporated. The outright refutation of the theory on the basis of the two-factor paradoxes of the U.S. and other countries is unjustified since the inclusion of natural resources tends to explain these paradoxes as well as the C-L intensity reversals. However, more research is needed on such topics as the nature of the production function and complementarity and/or substitutability of different factors as well as the role of the factors not treated in this paper for further clarification of the real status of the Heckscher-Ohlin theory.[16]

[16] Some efforts in this direction are already being made. For example, a recent empirical work of D. B. Keesing attempts to show that the U.S. exports are "research" intensive ("The Impact of Research and Development on United States Trade," unpublished, Feb. 10, 1966). A more comprehensive study which tests not only the "research" factor (incorporating direct and indirect research requirements) but also "education" requirements is being prepared by R. Baldwin, a preliminary result of which agrees with Keesing's ("Economic Characteristics of U.S. Exports and Import-Competing Industries," research in progress, Univ. of Wisconsin, Dept. of Economics).

ADAM SMITH'S THEORY OF JUSTICE, PRUDENCE, AND BENEFICENCE*

By William F. Campbell
Louisiana State University

Do we not owe the Growth of Wine
To the dry shabby crooked Vine?
Which, while its Shoots neglected stood,
Chok'd other Plants, and ran to Wood;
But blest us with its noble Fruit,
As soon as it was ty'd and cut:
So Vice is beneficial found,
When it's by Justice lopt and bound;[1]

Adam Smith in many ways is Mandeville without paradox. The core problem for both Smith and Mandeville is how to channel self-interest into socially beneficial manifestations. The two major restraints, prudence and justice, that Smith explains in the *Theory of Moral Sentiments* are also important in the *Wealth of Nations*. The third characteristic of the perfection of human nature, namely, proper beneficence, figures only marginally in the *Wealth of Nations*. The reason for this is that beneficence for Smith is the result rather than the cause of economic growth and development. As Smith says in the *Theory of Moral Sentiments*, "Before we can feel much for others, we must in some measure be at ease ourselves."[2]

Although there has been much discussion on the relationship between the *Theory of Moral Sentiments* and the *Wealth of Nations*, it has usually been carried on by partisans who uphold one book at the expense of the other. Unfortunately, little attempt has been made to carefully analyze the two books as complementary. When one does this, one notices certain relationships and ideas which have been obscured. First, Smith's formulation of justice is similar to certain problems of modern day welfare economics; second, the meaning of justice functions on two different levels: interpersonal relationships which are stressed in the *Theory of Moral Sentiments*, and intergroup relationships which are stressed in the *Wealth of Nations*. The impartial spectator becomes less of a spectator and more impartial when it is applied to economic policy as a rule of law norm. We shall see that much of Smith's distrust of governmental policy stems from his dedication to justice and impartiality rather than to an abstract devotion to free-

[1] Bernard Mandeville, *The Fable of the Bees*, F. B. Kaye, ed. (Oxford: Clarendon Press, 1924), Vol. I, pp. 36-37.
[2] Adam Smith, *The Theory of Moral Sentiments* (London: Cadell and Davies, 1812), p. 357.

dom. The natural law influences on Smith are more powerful and relevant than they have often been thought to be.

Before we can understand Smith's ideas on justice, we must examine the broader framework of his moral philosophy. George Stigler once observed that "the desire for better men, rather than for larger national incomes, was a main theme of the classical economics."[3] If this is true, what, then, constitutes better men for Smith? The following passage should illustrate what type of man Smith wished to develop: "And hence it is, that to feel much for others and little for ourselves, that to restrain our selfish, and to indulge our benevolent affections, constitutes the perfection of human nature; and can alone produce among mankind that harmony of sentiments and passions in which consists their whole grace and propriety."[4] This passage along with many others which could be cited should be sufficient evidence to indicate that Smith was aware of the possible abuses of self-interest. Smith never glorifies selfishness, greed, and an unbridled pursuit of personal gain, either in the *Moral Sentiments* or in the *Wealth of Nations*. It is the purpose of Smith's moral, legal, and economic thought to devise the appropriate institutional framework within which self-interest can be expressed without inflicting harm on other individuals.

We should recognize that Smith theorizes on two different moral levels. The first level is the one of propriety or the proper government and direction of all our affections. It is on this level that the truly moral individual lives. Propriety is not unrelated to the "golden Mean" in Aristotelian thought or even to justice in Plato's thought. All actions and affections of the heart can be judged in terms of whether it is proportionate to the cause which excites it. If we may use an analogy which Smith did not use, we may say that propriety is similar to a moral radar which is constantly indicating marginal adjustments in actions and sentiments. Although we do not intend to fully explicate this level of Smith's moral philosophy, it is important to recognize that the psychological process of sympathy and the construction of the impartial spectator are essential to the functioning of propriety. Furthermore, the method appropriate to this aspect of moral philosophy is the most concrete and empirical. To live on a fully moral level requires knowledge and appreciation of all those little circumstances which determine men's concrete actions in specific situations.

The second level on which Smith theorizes is more relevant to our purposes and will therefore be discussed in detail. On this second level, general rules can be formulated which is the highest level to which or-

³ George J. Stigler, *Five Lectures on Economic Problems* (London: Longmans, Green and Co., 1949), p. 4.
⁴ Smith, *Moral Sentiments*, p. 32.

dinary men can aspire. General rules are a second-best substitute for true moral behavior. They are devices by which insensitive and fallible human beings can have a rough measure of their actions, and by which men either through the force of public opinion or through the enforcement of the courts can control their antisocial tendencies and behavior.

More specifically, Smith says, "The man who acts according to the rules of perfect prudence, of strict justice, and of proper benevolence, may be said to be perfectly virtuous."[5] Let us consider these three virtues in turn. Prudence, for Smith, is the kind of cool, calculating behavior that we associate with enlightened self-interest. Prudence dictates the necessary conditions for preserving and increasing one's external fortune. In his description of the prudent man, Smith gives an example which illustrates his knowledge of human nature, and also illustrates that the academic world has not changed much since his time. He says with respect to the prudent man of letters: "For reputation in his profession he is naturally disposed to rely a good deal upon the solidity of his knowledge and abilities; and he does not always think of cultivating the favor of those little clubs and cabals, who, in the superior arts and sciences, so often erect themselves into the supreme judges of merit; and who make it their business to celebrate the talents and virtues of another, and to decry whatever can come into competition with them."[6] Prudence is, of course, also important in the *Wealth of Nations* as the source of the "desire to better our condition," and it is prudence which recognizes the advantages of the division of labor. Prudence is therefore the basis of saving and capital formation which is the engine of economic development for Smith. Even with all these advantages Smith has to say that prudence, though respectable and agreeable, is not one of the "most endearing, or of the most ennobling of the virtues. It commands a certain cold esteem, but seems not entitled to any very ardent love or admiration."[7]

Concern for the happiness of others recommends to us the virtues of justice and beneficence. Justice "restrains us from hurting" the happiness of others, and beneficence "prompt us to promote that happiness."[8] Smith prefers to use the term "justice" in its narrow meaning of abstaining from doing our neighbor any positive harm. In more familiar terms, it is the right to "life, liberty, and property." Justice is the limitation implied in the quip: my freedom extends as far as your nose. In fact, Smith's concept of justice is the necessary complement to freedom defined as the absence of coercion. Justice is the limit

[5] *Ibid.*, p. 418.
[6] *Ibid.*, p. 372.
[7] *Ibid.*, p. 377.
[8] *Ibid.*, p. 464.

or boundary beyond which no individual's pursuit of self-interest can extend. Furthermore, the observation of this limit is not to be left up to individual discretion as in the case of prudence (or as we also shall see of beneficence): it can be extorted by force.

In another passage Smith says that "beneficence is always free, it cannot be extorted by force. . . ."[9] Why does Smith make such a distinction between justice and beneficence? In modern terms it can perhaps be described as one way of distinguishing a free society from a good society. Justice is a necessary requisite for the existence of any civil society, and particularly a free society. Beneficence, although it is not a necessary condition for the free society, may be a necessary condition for a good society. Beneficence is an "ornament which embellishes, not the foundation which supports the building."[10]

For an appraisal of justice and natural law influences in Smith it is important to note that the rules of justice are compared to the rules of grammar, and the rules of the other virtues such as prudence and beneficence are compared to the rules which critics lay down for the attainment of what is sublime and elegant in composition. This is an important departure from the rationalist theorists of natural law by whom Smith was nurtured as a student—Grotius, Pufendorf, and the "never-to-be-forgotten Hutcheson." No longer is the analogy to mathematics or arithmetic. Without going into the philosophical difficulties, the rules of mathematics have an absoluteness and an a priori fixity, an independence from human will, that the rules of grammar do not have. Presumably grammar differs from society to society as the rules of mathematics do not. However, for any person living in a concrete society, the rules of grammar are fixed, and one can live with a long-run certainty that they will not be changed tomorrow. The rules of justice are established by a similar process; like grammar they are established by a slow process of social consensus, they are fixed for long periods of time without alteration, thus permitting individuals to make long-range plans without worrying about their being switched overnight.

The fact that the purpose of justice is to prevent one person from harming another is similar to problems of externalities where the actions of one producer or consumer affect adversely the utility or production function of another. Although Smith undoubtedly had in mind such obvious examples as murder, rape, and theft, it is instructive to interpret Smith's characterization of justice as a negative formulation of Pareto optimality. When one interprets Smith's formulation in this

[9] *Ibid.,* p. 131.
[10] *Ibid.,* p. 146.

manner, his discussion of the development of social institutions is useful to shed light on the difficulties of resolving externalities.

The Pareto rule which follows from Pareto optimality has been stated in the following manner: ". . . any social change is desirable which results in (1) everyone being better off or (2) someone being better off and no one being worse off than before the change."[11] For Smith, a change (an action) cannot be called socially beneficial if it involves injury or harm to other individuals. Such actions should be discouraged by means of the formulation of legal and moral rules. If one accepts this Pareto-rule interpretation of justice, then the problem becomes one of practical implementation. Little thought has been applied to how, in fact, one would go about making the Pareto rule (voluntary agreement) relevant to a wider range of social phenomena than purely private market choices. Perhaps realistic thought on the possibility of other kinds of social institutions has been hindered because of the polarization around the dichotomy of purely voluntary market choices and collective-coercive group decisions. It has been assumed that the only alternative to market processes is collective group decisions.

The unanimity in group decisions which is necessary in a strict interpretation of the Pareto rule implies a fully rational process with a determinate voting procedure by which individual choices are consciously expressed. If the solution of all externalities was to be achieved by voluntary agreement (unanimity), then the frequency of intimidation and blackmail would increase considerably. Consensus, on the other hand, is a much looser criterion of general agreement. Smith would have pointed out that consensus (but not unanimity) is a common fact of social life. Language, morals and laws (civil and criminal law, but not legislation) are important examples of consensus.

In the resolution of externalities the role of the judge as an impartial spectator is to discover the consensus which exists among rational men and to formulate it into rules of law. The role of the judge is not necessarily to read person's utility functions (although the role of judge as an impartial spectator using the psychological process of sympathy is in a better position to attempt this than the legislator), but he does have to decide what constitutes harm or injury. To do this he has to appeal to what is reasonable; i.e., to what the social consensus is among rational men. The problem of resolving interpersonal utility comparisons becomes one of constructing representative or impersonal utility functions by which conflict can be adjudicated.

[11] James Buchanan, "Positive Economics, Welfare Economics, and Political Economy," *J. of Law and Econ.*, Oct., 1959, p. 125.

The formulation of justice in the *Theory of Moral Sentiments* that we have been examining is a restraint on self-interest imposed to protect one person from harming another. Before proceeding to the extension which Smith makes of the concept in the *Wealth of Nations*, let us make clear that it is only an extension. Smith does not abandon the old concept of justice (i.e., refraining from harming another), but only applies it to new phenomena. In the *Wealth of Nations* justice takes on new dimensions imposed by the necessity of articulating a sound set of principles as the basis for social policy. Justice functions on an intergroup or interclass level as well as the interpersonal level. The concept of the spectator (involving the psychological process of sympathy) is appropriate to individual moral development, but not to social policy. The meaning appropriate to social policy is impartial treatment which implies equality before the law; no individual or group is to be awarded special privileges or forced to endure special restraints: "To hurt in any degree the interest of any one order of citizens for no other purpose but to promote that of some other, is evidently contrary to that justice and equality of treatment which the sovereign owes to all the different orders of his subjects."[12]

The elimination of preferences and restraints which is the "simple and obvious system of natural liberty and natural justice" is the main policy theme of the *Wealth of Nations*. If one had to choose a single passage from the *Wealth of Nations* which expresses its policy intent most adequately, it would not be the invisible hand passage, which is so persistently misunderstood, but the following passage: "All systems either of preference or of restraint, therefore, being thus completely taken away, the obvious and simple system of natural liberty establishes itself of its own accord. Every man, as long as he does not violate the laws of justice, is left perfectly free to pursue his own interest in his own way, and to bring both his industry and capital into competition with those of any other man, or order of men."[13] The system of natural liberty and natural justice can only be understood as the contradiction of the systems either of preference or of restraint. The whole point of Books III and IV of the *Wealth of Nations* is to understand the demands of impartiality in the social order. Smith was very explicit about this organization in the Introduction to the *Wealth of Nations*. He points out that "the policy of some nations has given extraordinary encouragement to the industry of the country; that of others to the industry of towns. Scarce any nation has dealt equally and impartially with every sort of industry."[14]

[12] Adam Smith, *The Wealth of Nations* (New York: The Modern Library, 1937), p. 618.
[13] *Ibid.*, p. 651.
[14] *Ibid.*, p. lix.

The natural course of economic development (corresponding to the system of natural liberty) was one where neither preferences were granted nor restraints imposed. The whole point of Book III is to compare the actual course of development dictated by the "policy of Europe" with the natural course of development. The point of Book IV is to examine the systems of political economy which arose to justify the policies of Europe. Mercantilism did not act impartially to the development of industry carried on in towns; it gave it special privilege. Physiocracy did not act impartially to the development of agriculture; it wished to give agriculture special privileges.

Smith's running attack on monopoly can also be approached from the point of view of the impartial spectator. He contrasts the natural price, or the price of free competition, with the monopoly price. The monopoly price (usually but not always) results from the grant of special privileges to firms, or special restraints on potential entrants. Smith says, "Such enhancements of the market price may last as long as the regulations of police which give occasion to them."[15] The main problem of monopoly for Smith was the elimination of legal barriers to entry. This was true in both the commodity markets and factor markets.

In the *Wealth of Nations,* Smith evidences a sympathy for the plight of the laborer. This approbation of the claims of the workingman is, however, the approbation of the impartial spectator; labor has been placed under special restraints, and owners have been granted special privileges. The answer to this problem is not to grant special privileges to labor, but to take away the special privileges of owners and manufacturers, and eliminate the legal restraints on labor.

Let us conclude by noting that we have been able to consider only one side of the interplay between moral development and economic development in Smith's thought. We have not been able to examine Smith's philosophy of history in which justice and beneficence are the fruits of an expanding commercial civilization rather than their cause. One aspect only must be noted. Smith was not guilty of the rather naïve Enlightenment views of the inevitability of progress. It is a tribute to Smith's realism and to his concern for the development of better men that he observed an important flaw in the very pillar of his economic scheme: the division of labor. Progress was not unilinear. The evils that were the inevitable accompaniment of economic progress should be ameliorated by prudent government action.

[15] *Ibid.,* p. 62.

THE EFFECT OF STATE FAIR EMPLOYMENT LAWS ON THE ECONOMIC POSITION OF NONWHITES*

By WILLIAM M. LANDES

University of Chicago

Introduction

Over the last two decades twenty-two states outside the South have passed laws prohibiting discrimination in employment on the grounds of race, creed, color, or national origin and have created commissions with powers to enforce these laws.[1] The orders of agencies administering these laws may be judicially enforced, and failure to comply can result in imprisonment and fines. During this period, many cities have established commissions on discrimination, generally without enforcement powers, designed to supplement the state commissions. Finally, the civil rights bill passed by Congress in 1964 includes an enforceable fair employment law of national scope.

Although legislators and civil rights groups have for a long time sought the enactment of new fair employment laws and the extension of those already on the books, little is actually known about the empirical effects of these laws in countering discriminatory employment practices.[2] Frequently, persons look upon the existence and continued passage of these laws as an indication of their success while, in fact, state fair employment laws only reveal a desire to reduce discrimination. Often, evaluations of the efficacy of these laws evolve into an enumeration of their enforcement provisions without considering the most basic question: Have fair employment laws improved the economic position of racial and religious minorities—especially the Negro minority? The purpose of my study is to investigate this question.

A theoretical model is first developed to analyze the impact of a fair employment law on the nonwhites' relative economic position. This is

[1] The twenty-two states with enforceable laws are New York and New Jersey (1945), Massachusetts (1946), Connecticut (1947), New Mexico, Oregon, Rhode Island and Washington (1949), Alaska (1953), Michigan, Minnesota, and Pennsylvania (1955), Colorado and Wisconsin (1957), California and Ohio (1959), Delaware (1960), Illinois, Kansas, and Missouri (1961), Hawaii and Indiana (1963). Employment discrimination is a misdemeanor in Idaho (1961), Iowa (1963) and Vermont (1963). For a summary of the provisions of these laws, see Paul H. Norgren and Sam E. Hill, *Toward Fair Employment* (Columbia Univ. Press, 1964), pp. 93–113.

[2] Two recent studies that attempt to assess the effectiveness of these laws are: (1) Norgren and Hill, *op. cit.*, and (2) Herbert Hill, "Twenty Years of State Fair Employment Practice Commissions: A Critical Analysis with Recommendations," *Buffalo Law Rev.*, Fall, 1964, pp. 22–69. Little quantitative evidence is presented to support the positions of the authors. Norgren and S. Hill argue that fair employment legislation has had some effect, particularly in New York. H. Hill claims that overall it has been a failure.

conveniently handled by a simple extension of the well-known utility maximization model under uncertainty where firms and unions decide whether to comply with or violate the law on the basis of which alternative yields a higher expected utility. The principal implication of the model is that a fair employment law, by increasing the cost of discrimination to employers and unions, should lead to an increase in the demand for nonwhites relative to whites, and hence an increase in the ratio of nonwhite to white wages and a fall in market discrimination against nonwhites. This hypothesis is then tested in the second part of the paper where empirical estimates are given for the effect of state laws on the ratio of nonwhite to white male earnings and market discrimination. In addition, evidence is also presented on the effects of fair employment laws on nonwhite-white unemployment differentials and occupational distributions.[3]

In the limited space of this paper I can do no more than highlight some of the theoretical and empirical results of my research on fair employment laws. Thus, in the theoretical section I skip many details and proofs, and instead stress the basic implications on the nonwhites' economic position when employer discrimination is outlawed. No analysis is provided for the law's impact on employee, union, and consumer discrimination. I can only state that the effect of the law on the nonwhites' economic position is generally independent of the source of discrimination. In the empirical section the more important quantitative effects of fair employment laws are merely stated without describing the technical problems encountered in making these estimates. Moreover, there is no discussion of my findings on the importance of differences among states in nonwhite-white schooling, the proportion of nonwhites, urbanization and tastes for discrimination as explanatory factors of observed differences in the nonwhites' relative economic position across states. I expect in the near future to publish a detailed study of the economics of fair employment laws which will remedy the gaps of this paper.

Theoretical Analysis

In states with enforceable fair employment laws it is unlawful for an employer to (1) refuse employment to a nonwhite because of his race (referred to as the segregation provision) or (2) discriminate in terms and compensation of employment against nonwhites for firms that employ both whites and nonwhites (referred to as the wage differential

[3] In effect, this is a study of Negro discrimination since more than 90 percent of persons classified as nonwhites in the United States are Negroes. Religious discrimination is not studied because virtually no relevant data can be obtained for religious minorities. Since racial discrimination is a more important issue and the primary target of fair employment legislation, this empirical restriction is unimportant in appraising these laws.

provision).[4] Provision (1) is necessarily ineffectual without provision (2). In the absence of (2) a firm would comply with the law if it offered a qualified nonwhite employment at a wage equal to zero or a small fraction of the white wage. Clearly, the nonwhite would choose not to work in this firm and hence compliance with the law would not have altered market behavior. For provision (1) to have any effect, there must be a minimum wage below which a firm cannot offer employment to a qualified nonwhite without violating the law. Provision (2) defines this minimum as equal to the white wage. Implicit in these provisions is their exclusive applicability to whites and nonwhites of equal economic capacity or marginal products. There is no violation when a firm refuses to hire a nonwhite less qualified than a white or pays a nonwhite a lower wage than a more productive white.

The impact of a fair employment law on employer discrimination can be treated as an added cost of discriminating to the firm. The magnitude of this cost equals the actual or realized cost of violating the fair employment statute, if one is caught (e.g., adverse publicity, costly litigation, and possible imprisonment and fines), times the probability of apprehension. This product is defined as the expected cost of violation. Specifically, we denote by $Ps \cdot Cs$ and $Pw \cdot Cw$ the expected costs of violating the segregation and wage differential provisions, respectively, for the firm per unit of time. Ps and Pw are the probabilities of apprehension while Cs and Cw are the realized costs or penalties if caught.[5]

Assume that white and nonwhite workers are perfect substitutes in production, have equal marginal products, and that there exist differences in tastes for discrimination among firms.[6] Firms can then be

[4] These two provisions, which are the key ones in all state statutes, apply more generally to persons discriminated against on grounds of race, creed, color, or national origin. Some states also prohibit discrimination because of age or sex. Fair employment laws apply not only to employers but also to labor unions and employment agencies. See P. Norgren and S. Hill, *op. cit.*, pp. 93–113 for a detailed account of the specific provisions of the state laws and a discussion of the differences among state laws.

[5] The concept of an expected cost of violation has wide applicability in interpreting behavior towards restrictive legislation. Illegal activities such as gambling and adultery are widespread, in part, because of the negligible probability of detection by the authorities rather than the magnitude of the penalty. Similarly, there may be widespread violation of fair employment statutes because of a small probability of detection rather than a low realized cost.

[6] The concept of a taste for discrimination or a discrimination coefficient is the crux of Gary Becker's development of a theory of discrimination. (See his monograph, *The Economics of Discrimination* [Univ. of Chicago Press, 1957].) Although I cannot summarize Becker's book in a paragraph, some elements of his theory must be presented because an understanding of the impact of fair employment laws presupposes a knowledge of Becker's theory of discrimination. Becker introduces discrimination coefficients for employers, employees and consumers, which transform money to net payments, and shows how these coefficients are manifested in the market through white-nonwhite earning differences. For example, an employer calculates the net wage of a nonwhite worker as the money wage plus a monetary increment equal to his discrimination coefficient multiplied by the nonwhite wage. The wage that is relevant to the employer in calculating the number of nonwhites to hire is the net and not the money wage of nonwhites. It follows that if whites and nonwhites have equal marginal products, the employer has a taste for discrimination and if white and nonwhite money wages are equal, the firm will hire only whites since their wage is less than the net wage of equally productive

divided into three groups depending on the magnitude of their tastes (i.e., discrimination coefficients). In the absence of a fair employment law, segregated nonwhite firms *(SN)* employ only nonwhites because each one of these firm's discrimination coefficient is more than compensated by the difference in white-nonwhite money wages. For *SN* firms

(1) $$Wn(1 + t_i) < Ww.$$

Integrated firms *(I)* employ both whites and nonwhites together as

(2) $$Wn(1 + t_j) = Ww,$$

and segregated white firms *(SW)* with

(3) $$Wn(1 + t_k) > Ww$$

employ only whites. *Wn* and *Ww* are the money wages of nonwhites and whites, respectively, and the *t*'s are the various firms' discrimination coefficients. $Wn(1+t_a)$ for $a=i, j, k$ is the net nonwhite wage which varies among firms depending on the size of their discrimination coefficient. From (1), (2) and (3) we observe that the largest t_i is less than t_j, and the smallest t_k is greater than t_j. All firms maximize net as distinguished from money income and, therefore, hire factors until their net wage equals the value of their marginal product.[7] Firms with the smallest tastes have the lowest net costs and tend to expand relative to other firms.

The behavior of *Sn* firms does not violate the law even though *Wn/Ww*, which is less than one, is partly determined by their discrimination coefficients. These firms cannot be accused of discrimination in employment against whites as they are willing to hire whites at the same money wage as nonwhites. Whites, however, refuse employment in these firms as money wages are higher elsewhere. Since *Sn* firms employ only nonwhites, the provision of the law regarding wage differentials within firms is, of course, not relevant.

---

nonwhites. To obtain employment nonwhites must offer to work at a lower money wage to compensate employers for their tastes for discrimination. When the white money wage equals the nonwhite net wage, the firm will be indifferent to whites and nonwhites and both will be employed. The white-nonwhite money wage differential is determined in the market in the following way. If the ratio of nonwhites to whites is small, nonwhites can find work with firms that have small discrimination coefficients. As a result, the ratio of nonwhite to white wages for perfect substitutes would approximate one. As the percentage of nonwhites rises, nonwhites must find employment in firms with progressively more intense tastes for discrimination, and these firms require a larger money wage differential to employ nonwhites. Hence, the relative nonwhite to white money wage will fall as the proportion of nonwhites increases and a negatively inclined relative demand curve is generated for perfectly substitutable whites and nonwhites. This demand curve together with the relative supply curve determines the white-nonwhite wage differential.

[7] It also follows that the monetary value of the nonwhite marginal product is more than the nonwhite money wage. This is equivalent to demonstrating that firms with a discrimination coefficient greater than zero do not maximize money income but instead must give up money income to satisfy their taste for discrimination.

The enactment of the law adds a cost to the set of SW firms for they systematically employ only whites, and this is prima facie evidence that they discriminate in employment against nonwhites.[8] As a consequence, these firms must choose among alternatives that can conveniently be specified by a simple utility maximization model under uncertainty. The choice between compliance and violation of the law is analogous to the utility maximizer choosing between gambling and insurance.

Assume firms maximize utility of net (not money) income and receive no utility or disutility from either compliance or violation of the law per se. We have $U = U(Y)$ and $U'(Y) > 0$ where U is the firm's utility function with continuous first and second order derivatives, Y is its net income and $U'(Y)$ is the first derivative of U with respect to a change in Y. A given SW firm that obeys the law has a utility of income with certainty equal to

$$(4) \qquad\qquad U^* = U(Y_1 - L^*)$$

where

$$(5) \qquad\qquad L^* = N(Wn^*(1 + t_k) - Ww)$$
$$= NWwt_k.$$

Y_1 is the firm's income before passage of the law, L^* is the certain loss from complying with the law, N is the number of nonwhites the firm hires to comply with the law, and Wn^* is the money wage the firm pays to nonwhites which for full compliance must equal Ww. L^* is clearly positive when Wn^* equals Ww.

Alternatively, the firm may choose to violate the fair employment law by continuing to refuse employment to nonwhites. In this case expected utility equals

$$(6) \qquad\qquad \overline{U}s = (1 - Ps) \cdot U(Y_1) + Ps \cdot U(Y_1 - Cs).$$

If Cs is less than L^*, the law would be ineffective because the loss from voluntary compliance would be more costly than losses resulting from apprehension. However, Cs must be greater than L^* since Cs includes not only L^* (for the firm caught must now comply with the law and employ nonwhites) but also possible fines, litigation, and the monetary equivalent of adverse publicity.

The choice between total compliance and noncompliance for SW firms that differ only with respect to the shape of their utility function is illustrated in Figure 1. Three utility functions are depicted: (1) firms with the function ABC are "risk avoiders," (2) firms with the function

---

[8] SW firms are more common than might be suggested from this definition. They include firms that hire nonwhites in menial jobs but refuse to employ qualified nonwhites (i.e., perfect substitutes for whites) in more skilled jobs.
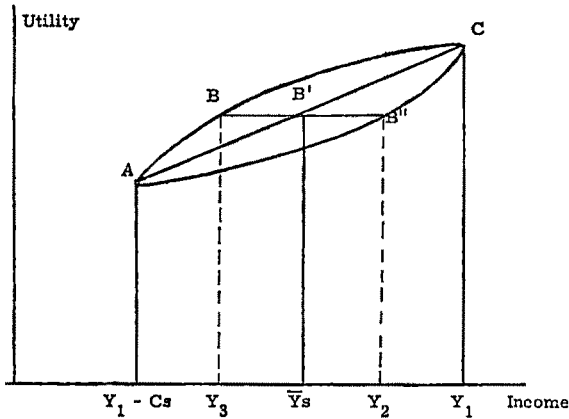
FIGURE 1

$AB''C$ are "risk takers," and (3) firms with the function $AB'C$ are indifferent to risk. The expected utility of an expected income is found for all three groups along the $AB'C$ chord. The distinguishing characteristic among the three groups is that given the choice between an expected income equal to an income with certainty, risk avoiders choose the latter, risk takers choose the former, and members of the third group are indifferent. Mathematically, the three groups may be distinguished by the second derivative of their respective utility functions $(U''(Y))$, which is negative for risk avoiders, positive for risk takers, and zero for the third group.

Risk avoiders have diminishing marginal utility of income and are willing to pay a positive premium up to $\overline{Y}s - Y_3$ to comply with the law, where $\overline{Y}s$ is the expected income from gambling with enforcement. If the premium exceeds $\overline{Y}s - Y_3$, the firm's income from compliance is less than Y_3 and it would, therefore, prefer noncompliance since the expected utility from the latter is greater than the utility from compliance. Thus, risk avoiders may sometimes find compliance too costly and prefer instead to risk enforcement penalties On the other hand, risk takers have increasing marginal utility of income and are willing to gamble with enforcement, hiring no nonwhites, provided the expected cost of this gamble is less than $Y_2 - \overline{Y}s$. For these firms the compliance premium has a negative value. However, if $Y_1 - L^*$ is larger than Y_2, risk takers prefer compliance as U^* is greater than $\overline{U}s$. Finally, firms with constant marginal utility of income have a zero compliance premium and violate the law only if $\overline{Y}s$ is greater than $Y_1 - L^*$.

Denote by π the compliance premium which equals the difference between $\overline{Y}s$ and the income from compliance that has the same utility as the expected utility of $\overline{Y}s$. The sign of π is positive, negative or zero

as $U''(Y)$ is negative, positive or zero. Therefore, utility maximization requires that SW firms employ whites and nonwhites together at equal wages, are indifferent to integration at equal wages, or continue to employ only whites according as

$$(7) \qquad\qquad \pi \underset{<}{\overset{>}{-}} \overline{Y}s - (Y_1 - L^*).$$

Generally, compliance will be most likely among firms with diminishing marginal utility of income and least likely among firms with increasing marginal utility of income, other things the same. For a given utility function the determinants of this decision are easily seen from (7). The greater the size of Ps and Cs, the greater the expected costs of violation, the smaller the expected income and utility from violation, and hence the more likely the firm will comply with the law. The smaller L^*, the greater the income with certainty and the more likely the firm will comply. L^* is positively related to the number of nonwhites the firm must hire to avoid possible penalties from enforcement, the size of the firm's discrimination coefficient, and the nonwhite money wage required for compliance. Thus, some firms with a t_k much larger than the market discrimination coefficient (t_j) may comply with the law due to a large $Ps \cdot Cs$.[9] However, if covariance $(Ps \cdot Cs, t_k)$ is less than or equal to zero, SW firms that comply will have a t_k much closer to t_j than do the set of all SW firms. We might expect SW firms to have a distribution of t_k's skewed to the right with a large fraction of these firms having a t_k close to t_j. This follows, for example, when the distribution of tastes among all firms approaches a normal distribution since the distribution of t_k excludes the lower tail of the taste distribution. If both the mean value of $Ps \cdot Cs$ and its variance are small, all firms may prefer to risk penalties rather than comply with the fair employment law.

A frequent criticism of state fair employment laws is that the hiring of one or two nonwhites by a previously segregated large white firm is sufficient to comply with the law.[10] "Token" compliance is explicitly handled by our analysis. For example, if N is small, the utility and income loss from compliance is small, and there is a greater likelihood that employers will choose not to gamble with enforcement. The few nonwhites employed in this manner are the main beneficiaries of the law as they receive an economic rent equal to $Wn^* - Wn$. However,

---

[9] The market discrimination coefficient can be computed from a knowledge of wages alone. From equation (2) it follows that $t_j = (Ww - Wn)/Wn$ for perfectly substitutable whites and nonwhites.

[10] See Herbert Hill, *op. cit.*, p. 38. Hill writes that "an employer may often feel his company is abiding by a nondiscriminatory policy if the Commission's poster is exhibited and one or two Negroes are employed." He adds that the real problem is "the fact that FEPC in Connecticut and elsewhere is not meant to change the Negroes occupational pattern by widespread and fundamental changes in racial employment practices."

maintenance of above equilibrium wage rates for a few nonwhites creates a rationing problem as it illicits a response from other nonwhites who have an incentive to try to enter the previously all white firm. A similar rationing problem would also exist in our previous analysis as even there the number of nonwhites receiving Wn^* is less than the number of qualified nonwhites. However, in the case where token compliance is sufficient to avoid the expected costs of violation, firms may openly reject nonwhite applicants after their small quota is filled.[11]

Before examining the direct effects of utility maximization by SW firms on the relative economic position of nonwhites to whites, let us first describe the response to the law by integrated (I) firms that in the absence of the law employ both whites and nonwhites at the market wage differential. I firms are now in violation of the wage differential provision. If this were the only restrictive provision of fair employment laws, these firms would avoid the law by refusing employment to nonwhites, which is the traditional interpretation of "equal pay" legislation. But in this case, I firms that dismiss nonwhite workers place themselves in violation of the "segregation" provision of the law. When this expected cost is added to white wages it is clear that total costs are a minimum for a given output when I firms employ only nonwhites at a money wage equal to Wn. In terms of Figure 1, I firms can continue to maintain an income of Y_1 with certainty, which yields the maximum utility, by employing only nonwhites.

We must further consider shifts in the supply of whites and nonwhites to I firms which result from the behavior of SW firms. Suppose some SW firms comply with the law, employing nonwhites at a wage of Wn^*. Nonwhites are attracted away from I (as well as SN) firms by higher wages in SW firms and, at the same time, SW firms will release some whites as they substitute nonwhites. Thus, the supply curve of nonwhites and whites to I firms will have shifted to the left and right, respectively. And as nonwhite wages rise and white wages fall to I firms it is no longer obvious that I firms will switch to all nonwhites. These firms are faced with a choice analogous to the one faced previously by SW firms. If Y_1 is the original pre-fair employment income, the hiring of only nonwhites at a wage equal to $Wn + \Delta Wn$ where ΔWn is positive will yield an income less than Y_1. On the other hand, employment of

---

[11] Various types of partial compliance for SW firms may be preferred to the alternatives of total compliance and noncompliance. For example, when token compliance is not entirely sufficient to avoid the law, some SW firms might still hire a small quota of nonwhites at wages equal to whites. This alternative would maximize utility provided the probability of apprehension or the realized costs were sufficiently lowered by this type of behavior. Another possibility is for an SW firm to hire nonwhites at a wage equal to Wn, which would violate provision (2) of the law. The limitation of space prevents me from discussing the effects of various types of partial compliance. I can only assert that the direction of their impact on the nonwhites' relative economic position is similar to that of full compliance.

whites only, at a wage of $Ww + \Delta Ww$, where ΔWw is negative, yields the prospect of an income Y_1' (which is greater than Y_1) with a probability $(1 - Ps)$ and an income $Y_1' - Cs$ with a probability Ps. I firms choose the alternative with the greatest utility. The determinants of this choice include the same factors that bear upon the decision of SW firms and, in addition, the changes in Wn and Ww. The smaller the changes in Wn and Ww the more likely I firms hire only nonwhites.[12]

The variables affecting the compliance decisions of SW, I, and SN firms have already been described. Since the empirical analysis investigates the effect of fair employment legislation on the relative economic position of nonwhites to whites, it is necessary to explicitly point out changes in the nonwhites' relative economic status that are suggested by the utility maximization model. SW firms that comply or partially comply will substitute nonwhites for whites. This will immediately raise the wage of some nonwhites. Since these nonwhites are attracted away from I and SN firms, the supply curve of nonwhites to I and SN firms will shift to the left, leading to an increase in nonwhite wages in these firms. Furthermore, SW firms release some whites which shifts to the right the supply curves of whites to I and SN firms. This will encourage some I and SN firms to substitute whites for nonwhites, dampening down the increase in nonwhite to white wages.

In sum, one predicts an increase in the "average" nonwhite to white wage ratio for a given proportion of nonwhites to whites. Thus a relative demand curve of nonwhites to whites would shift to the right. The more SW firms that comply, the greater will be the increase in nonwhite to white money wages. "Average" is added because all perfectly substitutable nonwhites need not receive the same wage after passage of the law. Nonwhites employed by SW firms receive a rent since their wage is higher than equally able nonwhites employed by other firms. Competitive market forces will produce pressures within the nonwhite group to eliminate wage differentials, and effective rationing would be required to maintain them. If, for example, the agency enforcing the law requires all firms to have a quota of nonwhites, then rationing of nonwhite employment in SW firms might be on a first-come-first-serve basis. Alternatively, complete compliance may require that all qualified nonwhites be hired. In this case, nonwhites would continue to come into SW firms until their wages were equalized in all firms. Should no firms prefer compliance, the effect on Wn and Ww will be slight as only those firms caught would pay a penalty that includes employment of nonwhites at wages equal to whites.

[12] These wage changes may be sufficient to induce some previously all nonwhite firms (SN) to employ only whites. This will be more likely when Ps and Cs are small, and when the firm's discrimination coefficient (t_i) is close to t_j.

Empirical Analysis

The quantitative effects of fair employment laws on the economic position of nonwhite males were estimated in a series of multiple regression models across states using readily available census data. The principal dependent variables included various measures of nonwhite-white male earnings, unemployment rates, and occupational distributions. The fair employment effects on these variables were isolated by holding constant differences among states in the proportion of nonwhite males in the labor force, the ratio of nonwhite to white male average school years completed, and urbanization. A dummy variable was inserted in the regressions to measure the fair employment effect. Several variants of the dummy variable were used to allow for differences among fair employment states in the number of years fair employment laws have been in effect and in the size of budgets of commissions administering these laws. Regressions were computed separately for all states in the U.S. and states outside the South. When the South was included in the analysis another dummy variable was added in order not to confuse fair employment effects with the effects of allegedly greater tastes for discrimination in the South. The major empirical findings may be summarized as follows.[13]

1. The ratio of nonwhite to white male wages (annual earnings divided by average weeks worked) tended to be higher by 5 percent and discrimination lower by between 11 and 15 percent in states with fair employment laws compared to states without these laws when differences in years of schooling, relative numbers, and urbanization were held constant. These results were obtained from a cross-section regression analysis in 1959.[14] The change in discrimination is measured by the percentage change in the market discrimination coefficient. I have previously indicated (see footnote 9) how the latter can be estimated from a knowledge of the ratio of nonwhite to white wages alone.

It is important to realize that the percentage change in discrimination is probably a better numerical indicator of the relative effectiveness of fair employment laws than the percentage change in the ratio of nonwhite to white wages. A hypothetical example will best illustrate this point. Suppose fair employment laws raise relative wages of perfect substitutes from .95 to 1.00. To argue that a 5 percent gain in relative wages is indicative of ineffectual laws is absurd. What is relevant is

---

[13] I have left out of this paper a number of interesting empirical, results that are secondary to measuring the effect of fair employment laws on the nonwhites' economic position. These include estimates of the expected costs of violation, a cost-benefit ratio, and variables influencing the passage of laws.

[14] Only states that passed fair employment laws before 1959 were counted as fair employment states for the empirical analysis. (See footnote 1.)

not the percentage change in relative wages but the fraction of discrimination eliminated. In this example, the law has been 100 percent effective, eliminating all observable discrimination.[15]

2. Based on an analysis of 1949 cross-sections, a higher level of relative annual income was not observed in that year for states that passed fair employment laws in the postwar period compared to states that did not pass laws. (Nine out of thirteen states with laws in 1959 passed their laws after 1949.) Thus, it is unlikely that relative wages were higher and discrimination lower in fair employment states previous to the enactment of this legislation. It was concluded, therefore, that the measured difference observed in 1959 could be associated with the passage of fair employment laws. This conclusion was further supported by an analysis of changes in relative nonwhite to white male annual incomes from 1949 to 1959 which showed that a significant increase in the percentage change in relative income took place in states that passed fair employment laws.

3. There are two reasons why the measured effect of the law on relative wages and discrimination is probably understated in the 1959 cross-section. First, the discrimination coefficient that is used as a base for the calculation of the percentage effect of the law is not net of differences in economic productivity between nonwhites and whites. Clearly, nonwhites have on the average lower marginal products than whites as evidenced, in part, by their fewer years of formal schooling and the inferior quality of this schooling. Thus, even if market discrimination were zero, the ratio of average nonwhite to white male wages in any state or region would be less than one. If, for example,

[15] One can show that the percentage change in discrimination as measured by the market discrimination coefficient (t) always exceeds the percentage change in relative wages for a small change in relative wages.

Proof: We have

$$t = \frac{Ww - Wn}{Wn} = (Wn/Ww)^{-1} - 1$$

and

$$\frac{dt}{d(Wn/Ww)} = -(Wn/Ww)^{-2}.$$

The elasticity of t with respect to Wn/Ww equals

$$E_t = \frac{dt}{d(Wn/Ww)} \cdot \frac{(Wn/Ww)}{t} = -\frac{1}{(Wn/Ww)t}.$$

Substituting $Ww/Wn - 1$ for t yields

$$E_t = -\frac{1}{1 - (Wn/Ww)}.$$

Since $0 < Wn/Ww < 1$, E_t must be less than -1. Hence a 1 percent increase in Wn/Ww leads to a greater than 1 percent decrease in t. Moreover, the nearer Wn/Ww is to 1, the greater is the percentage fall in t for a 1 percent increase in Wn/Ww.

average relative wages in the absence of discrimination equaled .850 in fair employment states, then these laws would have been associated with about a 25 percent reduction in market discrimination.[16]

Second, there is some evidence that the ratio of nonwhite to white male incomes was lower in 1949 in states that passed fair employment laws from 1945 to 1959 than in other non-southern states that did not enact laws in this period. This suggests that the cross-section estimates in 1959 understate the impact of these laws because they ignore the estimated negative difference found in 1949 and assume instead a zero difference in the absence of laws.

4. From 1939 to 1959 nonwhite-white unemployment differentials have grown larger in fair employment states relative to other states when one holds constant among states changes in the proportion of nonwhites, years of schooling of nonwhites relative to whites, and urbanization. However, in the initial period from 1939 to 1949 there was no systematic difference in changes in unemployment differentials between states with and without laws in 1959. In contrast, during the years after 1949, when the majority of states passed laws, there was a striking increase in unemployment differentials in fair employment states. Thus, instead of the ratio of nonwhite to white male unemployment rates rising on the average by 10 percent in fair employment states from 1949 to 1959, it would have fallen by about 10 percent had these states not passed laws. Similarly, the difference between non-white and white male unemployment rates would have fallen by more than 40 percent in the absence of laws instead of increasing by nearly 150 percent.[17]

A cross-section analysis of average weeks worked in 1959 gives additional evidence that fair employment laws have been associated with a widening of nonwhite-white unemployment differentials. It was estimated that the ratio of nonwhite to white male average weeks worked was nearly 3 percent lower in fair employment states compared to other states outside the South when differences among states in the proportion of nonwhites, nonwhite to white years of schooling, and urbanization were held constant.

The finding that fair employment laws have increased nonwhite-white unemployment differentials is not surprising. The theoretical analysis has pointed out that one consequence of the law is that non-whites employed by previously all white firms will earn an economic

[16] The ratio of nonwhite to white wages or marginal products in the absence of discrimination was approximated by estimating what relative wages would be if nonwhite and white wages differed only from differences in years of schooling and urbanization.

[17] In 1959 six of the thirteen fair employment states and seven of the nineteen non-southern non-fair employment states had state minimum wage laws. Thus, it is unlikely that the observed increase in relative wages and widening in unemployment differentials in fair employment states compared to states without these laws outside the South are attributable to state minimum wage legislation.

rent. In particular, a firm may comply with the law if it hires a few nonwhites at a money wage equal to the white wage but greater than the money wage of other equally able nonwhites.[18] This in turn would induce some nonwhites to leave their present job with the aim of finding employment in firms paying higher wages, or when involuntarily unemployed to search for a longer period of time in hope of finding a position yielding economic rent. The combined effect would increase the average length of unemployment per nonwhite and widen nonwhite-white unemployment differentials.[19]

5. The relative occupational distribution of nonwhite to white males has remained approximately unchanged in fair employment states relative to other states since 1939.

6. Tests were carried out for possible variations in the impact and enforcement of state laws due to differences in the number of years state laws have been operative and in the size of fair employment commission budgets. Generally, the allowance for these differences did not produce any significant effects on either wages, unemployment differentials or occupational distributions.

---

[18] The existence of rents would also mean an increase in observed income inequality among nonwhites in fair employment states. I expect to investigate this hypothesis in the near future.

[19] The law can have other effects on unemployment differentials but these will be transitory in that price adjustments will work to eliminate them. For example, the increase in demand for nonwhites relative to whites may lead to a reduction in nonwhite unemployment and an increase in white unemployment given a lag in the adjustment of relative wages to a shift in demand. Alternatively, some integrated firms that are violating the equal pay provision may choose to dismiss nonwhites, which would then violate the segregation provision, in preference to raising the wage of its nonwhite employees. This might be chosen if the expected cost of violating the wage differential provision was larger than for violating the segregation provision. There is some empirical support that a portion of firms do in fact choose the alternative of dismissing nonwhites. Of the 7,725 employment complaints filed with the New York State Commission from 1945 to 1962, roughly 25 percent charged employers with dismissing persons because of their race, religion, etc.

DISCUSSION

RICHARD R. NELSON: Naya's paper is an interesting and useful one. Before commenting on what I regard as its most useful conclusion, I want, first, to express a worry I have about its basic conceptual structure, and, second, provide a touch of balance to the natural resource explanation of the Leontief paradox.

There is something quite worrisome about the methodology, following Leontief, of calculating the direct plus indirect inputs which would be required if imported goods were to be produced domestically (and their inputs produced domestically too) and comparing this with the direct plus indirect input requirements of exports. While this would appear the correct approach for comparing the resource requirements of satisfying the existing final demand vector with existing trade patterns versus full autarchy, if the base of comparison is less than full autarchy, the conclusions can be quite misleading. Consider what would be required (say in Latin America) if a presently imported good (say automobiles) were to be produced domestically. The capital and labor for an automobile assembly plant certainly would be required. But there is no reason why the country would have to establish plant and allocate labor to the production of steel, or even of steel forms and shapes (the parts of the body or engine blocks). These could be imported. To say that a country saves both capital in the auto assembly industry and capital in the steel industry by importing automobiles is just wrong.

Nor, to get more directly to Professor Naya's paper, is there any reason to believe that a country that imports automobiles or airplanes is thereby conserving its natural resource availability of steel or aluminum, or that a country that exports these items is drawing on its natural resource availabilities. Canada exports aluminum and imports airplanes. Britain exports airplanes and imports aluminum. There is no reason why a country short on natural resources should not export products with a large natural resource content (direct and indirect) which it attains through imports. Whether it will or not will depend on such factors as transport cost of the raw materials versus the finished product, and its comparative advantage in the industry which uses the raw and semiprocessed materials as inputs.

Despite this problem, Kravis, Vanek, and now Naya have demonstrated the importance of natural resource endownments in explaining certain aspects of trade patterns. Certainly this is so for agriculture and mining products trade per se, and for trade in those products where it is much more costly to ship the raw materials inputs than the final product. But this does not explain U.S. exports of aircraft, computers, and machine tools, or Hong Kong's exports of textiles. Here I think the recent work of Keesing and the group associated with Vernon have just about nailed down the (long intuited) explanation: research and development, skills, and (possibly to a lesser extent) scale differentials among nations. This is not the time or place to review their work. Let me simply say here that research and development activity and skill re-

quirements in production both identified by Keesing as powerful explanatory variables behind the pattern of U.S. exports are strongly colinear. Peck, Kalachek, and I, in a forthcoming study, have observed and commented on this independently and in another context. We have argued that new processes and processes associated with new products tend to be both relatively unroutinized and in flux, thus requiring a flexible work force with redundant skills (since it is not known in advance what they will need to know or do) and precluding the use of intensive mechanization (which generally is possible only in routinized operations). As the product or process matures it becomes routinized, job specifications become more precise, and it becomes possible to substitute machinery for skilled labor and to train relatively less educated workers to perform most of the (now more routine and predictable) tasks. This would appear to explain both the high skill intensity and the low capital intensity of U.S. exports (which as Vernon points out are, to a considerable extent, new products).

For me, the most important empirical result reported in Naya's paper was the demonstration that, if agriculture and mining are eliminated, industry rankings of capital intensity are quite similar across countries (or at least between the U.S. and Japan). This is not surprising. One always suspected that the special circumstances played a far greater role in agriculture and mining, and the range of technology variation was greater, than was the case in manufacturing. But it is good to have the evidence.

Let me turn now briefly to Professor Campbell's paper. I found interesting the connection he tried to draw between Smith's concept of justice and the prescription of welfare economics to constrain activities which generate external diseconomies. He might have gone on to say that Smith's interest in beneficence and the public policy prescription of trying to encourage activities that generate external economies also would appear somewhat related.

Campbell raises an interesting point when he suggests (implicitly) that modern economists should follow Smith in paying more attention to voluntary compliance to norms, or moral rules, as a social control mechanism. When one thinks about it, it is clear that a tremendous amount of human behavior is so regulated rather than by the standard carrot, stick, and enforced constraint mechanisms stressed in modern economic theory. Most people do not steal, not because they feel they will get caught if they do (in most cases it is easy to get away with petty theft), but because they feel it is wrong. People do take valuable things they find to the lost and found, and not largely for hope of reward. People vote in elections even when this involves inconvenience and though it is extremely unlikely their vote will matter.

It is clear that "moral" behavior is tremendously important to society and to the economy. Think of the added costs of law enforcement if people were deterred from breaking laws only by fear of the cops. Think of how much more inefficient and costly the running of an economy would be if contracts and agreements were fulfilled only because of the coercive forces of society. Professional economists tend to sneer at the "business ethics" teachings of the Harvard Business School, but think of how difficult it would be to get an

economy to run efficiently if the only control instruments were money profit, money loss, and legally enforced constraints.

Campbell stresses Smith's emphasis that the main social benefit of economic progress is the conditions for a just and beneficent society. It also is clear that a culture which contains a strong moral ingredient of justice and beneficence can greatly facilitate economic progress.

ANTHONY M. TANG: Professor Naya's paper represents another significant contribution to a growing body of literature on the "scarce factor paradox" in the factor proportions theory of trade. To many of us, his conclusion that "empirical applications of the factor proportions theory will be more mean- ingful when factors in addition to capital and labor are explicitly incorpo- rated" might well appear familiar, if indeed not trivial. Yet, the literature is at times marked by an apparent unwillingness to leave behind the two-factor world. Minhas' strained theorizing and his still more vulnerable empirical foundation are a case in point. Where a third factor is explicitly introduced, the effort is sometimes marred by unwarranted generalization without recog- nizing the limitations of even a three-factor world. Thus, although the role of natural resources in the paradox was duly noted by Leontief to whom the "paradox" is attributed, explicit treatment of them has led Vanek to genera- lize the limited statistical association between these resources and capital. While Vanek's work sheds useful light on the problem and contributes to methodology, its value is compromised by the shaky empirical foundation for the alleged complementarity between resources and capital.

It is to these matters that Professor Naya addresses himself. In so doing, he considers the trade patterns of Japan, India, and Canada and the influence of natural resource adjustment on capital-labor ratios for exports and import substitutes. His finding points out important exceptions to Vanek's generaliza- tion and how these exceptions lead to inconsistent results among countries when natural resource adjustment is made. He is equally successful in laying bare the insufficient empirical basis for Minhas' rejection of the assumption of strong factor intensity.

In the latter connection, Naya goes over much the same ground covered by Professor David S. Ball in a recent publication.[1] And one might well wish that Naya had proceeded with his adjustment successively to highlight the crucial differences between the several natural resource subsectors, as Ball has done. It would have been well also for Naya to speculate more on reasons why sepa- ration of subsectors is warranted, both conceptually and statistically, again as Ball has done. Had he attempted to do so, he may well have discovered an added reason (not listed by Ball) why, as a crude dichotomy, agriculture is so very different from the other natural resource industries. I have in mind the overwhelming importance of international capital flows in the exploitation of natural resources, with the notable exception of agriculture. The presence of giant international firms in mining and petroleum extraction, for example,

---

[1] David Stafford Ball, "Factor-Intensity Reversals in International Comparison of Factor Costs and Factor Use," *J.P.E.*, Feb., 1966, pp. 77–80.

tends to submerge the differences in relative capital position that otherwise exist between countries. This follows since such firms have full access to the world's leading money markets and need not be governed by the capital cost relevant to the countries in which they operate. This may well be an important basis for Vanek's empirical generalization had he excluded agriculture from it. Although Naya shows only good judgment in limiting the scope of the substantive part of his work, the reader is nevertheless likely to be disappointed by the off-handed manner in which he treats the cumulative light already thrown on the scarce factor paradox by the several published studies on human capital by Leontief, Kravis, Keesing, and Kenen—to mention only a few. It should be clear now that any attempt to explain away the paradox is bound to leave important questions unanswered unless one goes beyond Naya's three-factor framework.

What I have to say about Professor Campbell's paper is not so much a comment as it is a footnote to his tour de force in the history of economic doctrine —an increasingly rare exploit among today's technically-inclined doctoral students. It would be useful to remind ourselves that Smith's writings are best understood against the background of the politico-economic order of his time as embodied in the then virulent system of political economy known as mercantilism. His stress on natural liberty and justice was a clear reaction against politically-granted privileges and "loaded" social policies. This reminder is useful because Smith's doctrines are less applicable to situations where problems are rooted in such considerations as "natural" market structure, externalities, indivisibilities, interdependence of economic decisions, price system imperfections, and disparities in social decisions as given by the sum of individual actions and by "socialized" choice making. Such considerations constitute the crucial elements in modern growth theory and policy prescription.

It would also have been useful to distinguish among Smith's doctrines those that are peculiar to the problems of his time from those that are capable of generalization. Among the latter are: (1) The fundamental proposition that justice and beneficence are the fruits of economic expansion, not its cause. (2) The qualifying proposition that while economic growth lays the groundwork for justice and beneficence the process of growth is accompanied by iniquities. The first suggests two additional insights: the limited scope offered by tinkering with an existing system with a view of coming closer to the Pareto optimum as a source of justice and beneficence, and a similarly limited scope offered by improving allocative efficiency as a source of growth. The second proposition reminds us of the socioeconomic problems rooted in the structural and locational attributes of growth so familiar to today's students of economic development. It is these problems that moved Professor T. W. Schultz to observe that economic development generates gains as well as losses and that while the gains are well diffused, the losses tend to fall on particular groups of individuals or geographical areas.[2] Finally, although it is well to say with Smith that all this calls for "prudent government action," it might be useful to

[2] T. W. Schultz, "A Policy to Redistribute Losses from Economic Progress" (a paper presented at a Labor Mobility Conference, Ames, Iowa, Nov., 1960).

distinguish further among socio-politico-economic milieus that offer not only differing propensities for corrective social policy but also differing needs for it.[3] In this connection, Smith's own times need to be carefully set apart from today's settings.

My comment on Professor Landes' contribution bears on its organization and methodology. The purpose of the paper is to gauge the effect of state fair employment laws on the economic status of the nonwhites in terms of their relative wage, unemployment rate, and occupational distribution. The paper is mainly concerned with an exposition of a model in two parts. The first part is a straightforward application of Gary Becker's seminal work on the economics of discrimination. This enables Landes to form three analytical categories of employers, the SN, I, and SW firms. The second part deals with the behavior of the SW (segregated white) firms, the object of fair employment legislation. These firms have the option of complying with the law or violating it. The former choice is analogous to buying insurance at a known cost, the latter to gambling. The model that suggests itself is that of utility maximization under uncertainty. In these terms, the SW firms may be classified as risk-takers, risk-avoiders, and firms indifferent to risk. These are useful analytical elements that Landes brings out. However, he makes surprisingly poor use of them in the paper (though perhaps not in the dissertation).

The analytical framework is clearly too elaborate in relation to the testable consequences Landes draws from it. It sheds no light on either the relative unemployment rate or the occupational distribution of the nonwhite. The only inference it supports is that the effectiveness of the law can be measured by the improvement in relative nonwhite wage. Such a limited consequence can be deduced by simply noting the fact that fair employment legislation seeks to influence market behavior by raising the cost of maintaining discrimination to the SW firms. The resulting shift in relative demand in favor of nonwhite labor must, other things being equal, lead to a corresponding adjustment in relative nonwhite wage. In these regards, Landes' paper violates the law of parsimony.

Because of his reliance on a single testable consequence, the power of the test is weak. The fundamental statistical finding—that the fair employment states showed greater improvement in relative nonwhite wage during 1949-59 than the other states—can be shown to be consistent with other hypotheses. One such hypothesis is that the observed differential improvement is due to skill-selective nonwhite migration to the fair employment states (either attracted by the congenial legal atmosphere or by fundamental economic forces for which the legislation happens to be a "proxy") while the law itself has had little or no effect on discrimination in employment.

The measure chosen to reflect relative nonwhite wage change is open to challenge. While Landes' argument for using the fraction of discrimination removed is cogent, it is one-sided in that a direct measure (absolute or relative)

[3] W. H. Nicholls, *Southern Tradition and Regional Progress* (Univ. of North Carolina Press, 1960), and "An Agricultural Surplus as a Factor in Economic Development," *J.P.E.*, Feb., 1963, pp. 1-29.

on the difference in nonwhite-white wage ratios before and after passage of legislation tells us something otherwise buried in the index used in the paper. And it would have been interesting to see the outcome under the other measure. After all, a discrimination elimination rate of 50 per cent as produced by moving from a 1949 nonwhite-white wage ratio of 0.5 to a 1959 ratio of 0.75 should be more meaningful than the same elimination rate when obtained by moving from 0.95 to 0.975. One would also have thought that the elaborate model might have yielded useful insights into the attributes of the firms with varying risk-taking propensities. I have in mind, among other things, firm size.

Landes' investigation is interesting, timely, and imaginative. It is a matter of regret that the paper's organizational and structural flaws have crowded out what should be the most interesting part of the study; i.e., the empirical section. An equally unfortunate omission is the failure to summarize the enforcement section of state fair employment acts. The reader would no doubt find himself in a better position to evaluate Landes' findings when equipped with some notion how the law works.

DAS KAPITAL: A CENTENARY APPRECIATION
INTRODUCTORY COMMENTS

By JOHN M. LETICHE

University of California, Berkeley

Neils Bohr once defined a "deep truth" as a statement the opposite of which also contains deep truth.[1] According to this dictum, Karl Marx must have discovered many a deep truth.

To Joseph A. Schumpeter, the greatness of the Marxian conception might well dazzle the modern disciple. Marx's analysis, he insisted, was the only genuinely evolutionary economic theory that the period produced. Development was its central theme. Nobody, at that time, had a fuller appreciation of the size and power of the capitalist engine of the future. Although exaggerated emphasis on exploitation endangered the explanatory value of Marx's scheme, he recognized that for the social process as a whole the essential point of economic development was the capacity-creating use made of capitalists' gains.

Schumpeter considered neither Marx's assumptions nor his techniques above serious objections. He regarded the individual features of Marxian thought to be based upon a diagnosis of the social situation which was ideologically vitiated in its roots, hopelessly wrong in its prophecy of ever increasing mass misery, and inadequately substantiated both factually and analytically. Nonetheless, he considered Marx's performance "the most powerful of all."[2] Marx concentrated his analytical powers on the task of showing how the economic process, changing by virtue of its own inherent logic, incessantly changes the social framework and, in effect, the entire society. It was this grand vision of an immanent evolution, wrote Schumpeter, "that constitutes Marx's claim to greatness as an economic analyst."[3]

Similarly, Professor Bronfenbrenner, in the article prepared for this hundredth anniversary of *Das Kapital,* heralds Marx as "the greatest social scientist of all time," although "not necessarily the greatest economist." Professor Samuelson, on the other hand, deems Marx's contribution "from the viewpoint of pure theory, to be that of a minor post-Ricardian," albeit "a not uninteresting precursor of Leontief's input-output." He also reminds us that Lord Keynes alluded to *Das Kapital* "as an obsolete textbook . . . without interest or application for the

---

[1] Neils Bohr, *Atomic Physics and Human Knowledge* (John Wiley and Sons, 1958), p. 66.
[2] Joseph A. Schumpeter, *History of Economic Analysis* (Oxford Univ. Press, 1954), p. 573.
[3] *Ibid.,* p. 441.

modern world." Manifestly, according to Bohr's definition, Marx's writings contained many "deep truths"!

Bohr, however, was balancing seriousness with humor. Young physicists, he reports, came from many countries to work at his Institute in Copenhagen. When they reached an impasse, they comforted themselves with the jest that there are two kinds of truth. To one belong statements so simple and clear that the opposite assertation obviously could not be defended; to the other, so-called "deep truths," presumably belong statements to which the opposite had also seemed true. The development of knowledge, Bohr noted, usually passes through such stages in which chaos becomes gradually replaced by order. He spoke with trepidation about the importance of the utmost caution in all questions of terminology and dialectics, warning that in bringing order into an entirely new field of experience, we could hardly trust in any accustomed principles, however broad, apart from the demand of avoiding logical inconsistencies. The key issue, he maintained, was the endeavor to clarify the conditions, in each field of knowledge, for the analysis and synthesis of experience.[4]

All scientific knowledge—including economics—presents itself within a conceptual framework adapted to account for previous experience. Any particular framework is likely to prove too narrow for the comprehension of new experiences. Time and again it has been the widening of the conceptual framework which has disclosed false premises previously hidden and thus protected from scrutiny, as well as analogies with respect to analysis and synthesis of experience in apparently separated domains, suggesting the possibility of a more productive general description.

Discussions of Marxian and post-Marxian economics, comparable to the history of Keynesian analysis, are likely to yield greater returns if conducted in such nonideological terms. As the exploration of new fields of experience in physics has disclosed unsuspected presuppositions for the "application of some of our most elementary concepts and thereby given us an epistemological lesson with bearings on problems far beyond the domain of physical science,"[5] the exploration of new fields of experience in economics may disclose unsuspected application of some of our economic concepts both to market and to centrally planned economies. In this regard, the coexistence of the different intellectual traditions may prove to be indispensable to all other forms of coexistence. Clearly, it is in this spirit of scientific advance that the papers and commentaries for the centenary of *Das Kapital* have been formulated.

---

[4] Cf. particularly Bohr's essay on the "Unity of Knowledge," *op. cit.*, pp. 67-82, and the attention he draws to complementary features of analysis in various disciplines, pp. 62-66.

[5] Bohr, *op. cit.*, pp. 68-69.

NOTES ON MARXIAN MODEL OF
CAPITAL ACCUMULATION*

By ALEXANDER ERLICH

Columbia University

Economists are in the habit of being serious but not overly solemn in our meetings, which is wholly to the good. Allow me nevertheless to depart for a moment from the unwritten rule and to begin by saying that I consider it a very great privilege to be on the panel of this particular session. My reasons for feeling this way will, I hope, become clear from the presentation that follows.

An explanation of the slight change in the title seems in order. The notion of a model is associated these days with application of a more or less high-powered mathematics, and it is undoubtedly true that Marxian analysis of capital accumulation contained in Chapter XXV of the Volume I of *Das Kapital* cries out for a rigorous mathematical going-over. The same goes for the closely related "scheme of the expanded reproduction" presented in the concluding part of the Volume II and cast in the form of cumbersome arithmetics. But while the cry is duly echoed here, it will not be heeded: this writer's mathematics is much too featherweight for that. (My sense of regret on this score, let me add, is tempered by the realization that high mathematical skills are by no means in short supply in the profession, most obviously not among the distinguished group behind this table and that significant progress toward meeting the need I referred to has already been made.[1]) To call the present paper "Notes on Marxian Model" would be more in keeping with honest labeling practices.

The outlines of the story are familiar, and hence need be recapitulated only briefly. The accumulation of capital propelled by the individual capitalist's desire to increase his wealth and power over people and enforced as a condition of survival by the exigencies of the competitive struggle is bound sooner or later to outpace the increase in supply of labor; as a result, the wages rise and profits fall. The capitalists re-

spond by introducing laborsaving machinery which makes part of the work force redundant. Yet while the continuing interplay between accumulation and technological progress arrests and reverses the fall in the rate of exploitation, it opens up other roads to ultimate self-destruction of the system. The rise in amount of capital per worker reduces the rate of profit; progressing "immiseration" makes the aggregate effective demand lag behind the rapidly expanding productive capacity; and these intertwined processes which work their way through increasingly violent ups and downs of the business cycles set the stage for the victorious proletarian revolution. A full-scale discussion of the theories underlying this panoramic view can obviously not be attempted here. I shall, instead, focus on a few selected issues, and argue (1) that several crucial propositions put forward by Marx could stand up much better if recast in terms of modern economics; (2) that his model of capital accumulation was more complex and capable of a wider range of solutions than is generally assumed; and (3) that a tighter integration of this model with the analysis of Volume II would help in plugging some of the gaps.

I

Regarding the first point, I have very little to add to what was already said by Professor Joan Robinson in the path-breaking *Essay on Marxian Economics* [17] and elaborated on in her later writings. To begin with, Marx was clearly handicapped by his "Say's law" approach to the saving-investment problem, even though he rejected the notion that supply creates its own demand and noted that an injection of additional investment expenditure into a system in equilibrium could set off a boom by its demand generating effects. But he clearly sided with his contemporaries, not merely in assuming that capitalists' savings were made for the purpose of further investment, but also in taking for granted that these savings would materialize as intended and that they would in fact be reinvested after having been made. ("Therefore save, save, i.e., reconvert the greatest possible portion of surplus value or surplus product into capital!" [11, p. 652].) By so doing, Marx effectively knocked out the props from under his argument about limited "consuming power of the society" as the ultimate cause of insufficient aggregate demand.[2] But even if he had succeeded in cutting himself loose from Say also on this score, he would still need a force which could keep the volume of investment below the size needed to bridge

[2] It is true that Marx makes an attempt to base his underconsumptionist position on the alleged fact that "the production of constant capital never takes place for its own sake, but solely because more of this capital is needed in those spheres of production, whose products pass into individual consumption" [13, p. 359]—a rather weak and highly "unMarxian" argument.

the increasing gap between the productive potential of the economy and the level of total consumption. On the face of it, the famous "law of the falling rate of profit" seemed to be the right candidate for the job. Yet this law was a very weak rod to lean upon, and here too, one can hardly do better than follow the criticisms of Professor Robinson, joined on this particular issue by Dr. Paul M. Sweezy whose *Theory of Capitalist Development* [21] appeared simultaneously with her *Essay*. (A discernible note of doubt can be found already in Mr. Maurice Dobb's *Political Economy and Capitalism* [3], published in 1937.) As both of them pointed out, whether the rate of profit falls, rises, or stays constant depends on relative changes in the "organic composition of capital," and in the "rate of exploitation," and there are no compelling economic reasons why the first should increase more rapidly than the second.[3] Furthermore, technological progress need not always entail an increase in the organic composition of capital, particularly if capital is measured in terms of labor hours rather than in terms of physical volume. It might be added that in situations when an increasing amount of capital per man would go together with a decline or with constancy of the capital-output ratio, the likelihood of rising (or constant) rate of profit would be greater than in case of increasing capital-output ratio. Actually, Marx was candid enough to list explicitly these points (as well as a few others) under the heading of "counteracting causes." But his attempts to demonstrate that the law would nevertheless assert itself in the end were not convincing, and the same must be said about efforts made by several of his orthodox followers.[4]

Hence, in order to rescue the law from disintegrating, some variety of the "bourgeois" notion of diminishing returns to capital and/or declining marginal efficiency of investment would have to be brought in —indeed, as will be argued later on, Marx did come close to developing the last-mentioned concept. Similarly, in order to present "the limited consuming power of the society" as a factor of economic instability and as an impediment for growth, injection of the Keynesian distinction between intended and actual savings was indispensable. Barring such emendations and with the considerations of equity shunted aside as value judgments, the somber picture outlined in the introductory paragraph would turn into its opposite, to borrow one of

---

[3] Cf. Robinson [17, pp. 35–42], and Sweezy [21, pp. 100–08]. It may be worth noting here that Marx, in a little-known passage, tried to supply such an economic reason by taking recourse to a Ricardian-sounding argument. The value of labor power, he insisted, declines more slowly than it would correspond to the overall increase in productivity of labor, because increase in productivity in industry is faster than in agriculture which produces the workers' means of subsistence. Cf. [15, pp. 359–60].

[4] Cf. Rosdolsky [18]. The most elaborate and sophisticated defense of the orthodox Marxian position known to this writer was presented by Shane H. Mage in his unpublished Ph.D. dissertation [10].

Marx's favorite expressions. It would now become a vision of an economy whose savings are religiously plowed back into expansion and incessant technological improvement, with the major stop such as law of diminishing profitability pulled out for all practical intents and purposes and with the industrial reserve army keeping the labor supply elastic. A more eloquent eulogy for capitalism as a uniquely powerful and wear-resistent engine of growth would be difficult to conceive. It was not surprising that Schumpeter was so fond of invoking Marx against Keynes or that Tugan Baranovski was using tools from the armory of *Das Kapital* to develop his own version of J. B. Clark's "more mills that should make more mills forever," with consumption becoming increasingly expendable. (At this point Rudolf Hilferding, one of the leading theorists of the pre-1914 Austro-Marxism, commented, "If this is madness, there is a method in it, in fact—a Marxist method.")

II

In the preceding paragraphs we asked whether the assumption of steadily rising rate of exploitation could sustain the burden of one of Marx's two major prophecies and whether it was compatible with the other. It is not our task to inquire whether or not this assumption has been borne out by the actual developments in capitalist countries of the West since Marx wrote: we do know the answer. More to the point in the present context, and less easy to resolve, is another question: to what extent does the perspective of the "immiseration" (implying either an outright decline in the living standards of the workers or at least a pronounced decline in the relative share of national income going to them) inescapably follow from Marx's own premises? The tenor of the crucial Chapter XXV of the Volume I, and quite definitely of the often quoted passage near the end of the Section 4, seems to leave little room for doubt. But the analysis, if examined in its familiar as well as not-so-familiar aspects, tells a somewhat different story.

To begin with, we find that the race between accumulation and labor force can conceivably be resolved in three ways rather than one, depending on circumstances. Indeed, Marx explicitly envisages a situation when "the price of labor keeps on rising because its rise does not interfere with the progress of accumulation." There is, we are told, "nothing wonderful" about it; already Adam Smith knew that "a great stock, though with small profits, generally increases faster than a small stock with great profits" [11, pp. 678-79]; the sentence quoted from Adam Smith is found in [20, p. 93]. The argument seems awkward; the cited statement could not possibly be true with regard to the behavior of a capital stock over several successive time periods. It could apply, however, either to different capital stocks of widely ranging sizes at the same point of time, or to a single capital stock observed

at widely disparate stages in its life history, provided that this capital stock has been accumulating at an increasing rate at least over a considerable part of the intervening period. The last mentioned possibility would fit perfectly well the case of the opening up of a new area which Adam Smith was discussing when he made his dictum about stocks and profits. The increasing returns to capital resulting from such a situation could be expected to swamp the effect of rising wages, to be sure, only as long as these returns would persist. This was, presumably, what Marx had in mind when he approvingly cited Adam Smith, but he brought in another and more generalized explanation when he observed that "under special stimulus to enrichment, such as opening of new markets or of the spheres for the outlay of capital in consequence of newly developed social wants, etc., the scale of accumulation may be suddenly extended" [11, p. 672].[5]

The second alternative is less favorable to workers, and also more clearly defined. The wage increases eat into profits and thus reduce the volume of resources available for accumulation during the subsequent period.[6] As a result, the expansionary trend reverses itself and the demand for labor begins to taper off until the reduced rate of accumulation will have caught up with the rate of increase in the labor supply. However, even in this case Marx is cautiously noncommital in appraising the final outcome; "the price of labor falls again to a level corresponding with the need of the self-expansion of capital *whether the level be below, the same, or above the one which was normal before the rise in wages took place.*" ([11, p. 679]. Italics supplied.)

No doubt, after all is said and done, the third alternative—the shift to labor-displacing technology in response to increasing relative scarcity of labor—is both the best known and by far the most important.

---

[5] It undoubtedly seems odd that Marx allows accumulation to outrun the labor supply, although fixed capital-labor coefficients are assumed throughout—a point stressed by Professor Samuelson [19, p. 901]. As will be shown later, in a different context Marx explicitly admitted the possibility of increase in the number of workers without a concomitant increase in the volume of plant operated by them, but not the other way round. Furthermore, the assumption that a developing economy would not experience any technological progress whatsoever is, of course, extremely drastic even on the first-approximation stage; and it certainly helps to discard the most simple explanation of an increase in real wages which "would not interfere with the progress of accumulation," unless the technological progress is taken to be heavily and uniformly capital-using.

[6] As Professor Robinson rightly observed [17, p. 27] there is an interesting dualism in Marx's explanation of the mechanism behind the downturn in accumulation: he ascribes it at first to the weakened inducement to invest ("the stimulus to gain is blunted"—[11, p. 679]) only to return right afterward to the decline in volume of investible resources resulting from wage increases as the crucial factor. Indeed, the whole discussion has a distinctly pre-Keynesian flavor: divergencies in the rates of increase of capital and labor supplies are assumed to set off quasi-automatically a chain of equilibrating adjustments, with aggregate demand considerations left to look after themselves, and changes in inducement to invest being no more than a faithful reflection of the changes in the profit-wage ratio. True, also here the profitability schedule could be occasionally jerked upward as a result of "opening up of new markets or of the spheres for the outlay for capital," but in context of this particular discussion Marx clearly treated such a possibility not merely as an exogenous, but as a fairly incidental factor.

Yet here, too, the case is not as clear cut as it seems at first sight. Although Marx scathingly rejected the "compensation theories" of his contemporaries who believed in a quasi-automatic reabsorption of the displaced, he did outline several possible offsets of quite a different kind.

The first of them would probably not be considered a bona fide offset by Marx, and it is definitely least appealing (as well as most "static") of the lot. "Crippled as they [the displaced workers] are by the division of labor, these poor devils . . . cannot find admission into any industries, except a few of inferior kind, that are over-supplied with underpaid labor" [11, p. 481]. At a different place, but in a similar context, Marx spoke of "greater resistance which some lines of production, by their nature, put up against a transformation of *manofactory* into machine-operated production" and which enables them to make use of some of the "disposable or unemployed wage laborers" [13, p. 277]. To a present-day reader this looks like a description of the "dual economy" in action, very much in the spirit of Professor Eckaus' memorable box diagram showing how the expansion of capital-intensive sector can cause further lowering of capital-labor ratio in the labor-intensive sector. An alternative way of reading the quoted statements, particularly the first, would be to interpret them as a reference to "disguised unemployment" [17, p. 38]. In such case, obviously enough, the whole operation would not constitute any offset whatsoever.

"Although machinery necessarily throws men out of work in those industries in which it is introduced, yet it may, notwithstanding this, bring about an increase of employment in other industries" [11, p. 483]. The analysis that supports this proposition is very "dynamic" and highly modern in substance, although not in language. A distinction is drawn between resource-saving and output-increasing innovations, with a clear implication that the latter are predominant[7]; and the stimulus imparted by such an industry A to the complementary industries B, C . . . is traced in a way reminiscent of Professor Hirschman's backward-forward linkages.[8] Lastly, the impact of the economies of scale, working through a "feedback" type of repercussion, is shown to have powerfully affected, at an early stage, the very industry

---

[7] "If the total quantity of the article produced by machinery, be equal to the total quantity of the article previously produced by a handicraft or by manofactory, and now made by machinery, then the total labor expended is diminished. . . . But, as a matter of fact, the total quantity of the article produced by machinery with a diminished number of workmen, instead of remaining equal to, by far exceeds the total quantity of the hand-made article that has been displaced" [11, pp. 483-84].

[8] "As the use of machinery extends in a given industry, the immediate effect is to increase production in the other industries that furnish the first with means of production. . . . When machinery is applied to any of the preliminary or intermediate stages through which the subject of labor has to pass on its way to completion, there is an increased yield of material in those stages, and simultaneously an increased demand for labor in the handi-crafts or manufactures supplied by the produce of the machines" [11, pp. 484-85].

that was the natural vehicle of change: "As inventions increased in number and demand for newly discovered machines grew larger, the machine-making industry split up, more and more, into independent branches" [11, p. 417]. Such an impact, it would seem, could temporarily swamp the potential cyclical pattern built into the modern technology through growing importance of durable fixed plant as compared with "goods in process"—a point Marx was very explicit about.[9]

All of these "external effects" of laborsaving innovations could, of course, work themselves out only through investment. They clearly constituted one major category of the offset possibilities which Marx had in mind when he insisted that the displaced workers can be reemployed "only by an intermediary of a new and additional capital that is seeking investment, not at all by the intermediary of the capital that formerly employed them and was afterward converted into machinery" [11, p. 481]. Another batch of opportunities for new investment that could perform the same function was provided by technological changes which were in the nature of product innovations rather than process innovations: "entirely new branches of production, creating new fields of labor, are also formed, either directly on the basis of machinery, or at least in the wake of the general industrial revolution brought about by it" ([11, p. 487]; translation slightly corrected). Lastly, to the extent that the stringency of the labor supply had been lifted, there would now be scope also for "accumulation as a simple extension of production, on a given technical basis," even though the "intermediate pauses" during which such accumulation of a "widening" variety could occur would tend to become shorter in the long run [11, p. 690].

All this raises a host of points of interpretation which unfortunately cannot be pursued here.[10] But one thing seems quite certain: the no-

---

[9] "Further, the machinery need not be renewed till it is worn out. Hence, in order to keep the increased number of mechanics in constant employment, one carpet manufacturer after another must displace workmen by machines" [11, p. 479]. Cf. also the often quoted passage on the "machine builder" in [16, p. 355].

[10] A few of these points may nevertheless be briefly mentioned: (1) For economists brought up in the neoclassical tradition the notion of a jump from a labor-shortage to a labor-surplus situation as a result of "induced" laborsaving innovations seems puzzling. Shouldn't a shift to more capital using processes merely restore the equilibrium between supply and demand on the labor market? One way of answering this query would consist in noting that in the Marxian scheme of things the capitalists have every reason to carry the substitution of capital for labor beyond the point of full employment equilibrium; otherwise real wages could not be maintained at the level of reproduction costs of the labor power and would tend to eat into the surplus value. In short, "industrial reserve army" performs the same yeoman's service for Marx as "Malthusian devil" does for Ricardo. Yet while this argument, outlined in an early article by Oskar Lange [7] and later adopted by Sweezy [21], is not likely to cut much ice with economists who are not committed to the labor theory of value, Marxian analysis of technological change contains assumptions which in no way depend on such a commitment and which lend plausibility to the "overshot" thesis: (a) The technological progress in the case at hand is of a sharply discontinuous kind which makes the initial displacement effect on labor and on old-type

tions of steady and progressing displacement of labor by machinery and of the "accumulation of misery" as its concomitant turn out to be much less firmly anchored in the logic of Marx's argument than his own conclusions could lead us to assume. (No doubt, he could still argue that without the intervention of the laborsaving technological progress real wages would have amounted to a larger and rising fraction of the social product; but this would not do.) They emerge as a distinct possibility—no less, but also no more. The extent to which such a possibility might materialize would depend entirely on the volume of the "additional and new capital" that would be forthcoming. Actually, Marx came very close to saying this in so many words when he stated that "the invention of machines and application of the forces of nature sets free capital and people . . . unless new spheres of production will be created or the old ones will be expanded and operated on a much higher level" [14, p. 342]. He seemed quite sanguine about potentialities for such growth when he observed that "with accumulation, and the development of the productiveness of labor that accompanies it, the power of sudden expansion of capital grows also . . . because the technical conditions of the process of production themselves —machinery, means of transport, etc.—now admit of the rapidest transformation of masses of surplus product into additional means of production" [11, pp. 693-94]. However, this was only one part of the story he had to tell; and it is at this juncture that the analysis of Volume II must be brought in to round out the picture.

---

capital quite pronounced; at the same time this progress involves a dramatic increase in amount of (new-type) capital per worker in an economy that is not very well adept at supplying it—a point to which we shall return; (b) the capital-labor ratios within each particular technological method are essentially fixed which makes it harder to reabsorb the displaced; and although Marx admits that the degree of utilization of plant can be varied [11, pp. 661-62], the portent of this relaxation is limited, because the capital stock of the economy (and, a fortiori, of its nascent modern sector) is implicitly assumed to be too small to provide jobs for the whole "industrial reserve army" even if utilized to the full capacity; (c) the preceding propositions explain the emphasis put on new investment as the only relevant offset to replacement. The argument runs exclusively in terms of capacity-increasing effect of such investment; Marx gave no indication of recognizing its income-generating effect in this particular context, and insofar he underestimated the reabsorption possibilities. Yet here, too, the lack of significant capacity reserves must be borne in mind. Owing to the relatively small size of the capital stock, the short-run impact of the multiplier on output and employment would be limited, although its effect in "enforcing" profits would be significant. (2) Are we to assume that new investment Marx is talking about in the paragraphs quoted above would, in his view, not have taken place at all if the innovation had not occurred or that it would have found its way into the system in any case albeit at a lower rate of return? It seems that with regard to the major portion of this investment the first interpretation would be correct. As we saw, Marx assumed that the scale of accumulation would be "suddenly extended" (i.e., that saving and investment out of profits would increase at the expense of capitalists' consumption) "under special stimulus to enrichment such as the opening of new markets or of new spheres for the outlay of capital." In case of a less dramatic increase in investment opportunities such a "sudden extension" would presumably not occur. But also then it would still be true that "the additional capitals formed in the course of normal accumulation serve *mainly* as vehicles for the exploitation of new inventions and discoveries" ([11, p. 689]; italics supplied); and the same would hold for the full-scale replacement of the old capital.

III

The last two decades brought a very pronounced increase of interest in and of familiarity with Marxian schemes of "expanded reproduction." I shall therefore forego the detailed exposition and concentrate not on what the model of the Volume II is but on what it does and, more particularly, on the way in which it can help out with problems raised in preceding sections.

To put it in a nutshell, Marx's major feat consisted (1) in slicing the national income aggregate along a dividing line that was crucial for determining the growth potential as well as for keeping track of both contestants in the accumulation-labor race; and (2) in explicitly relating the output flows of the two major sectors, thus derived, to each other *and to capital stocks that produced them.* It is the feature indicated in the italicized part of the last sentence which is of particular importance to us. By translating the "reproduction scheme" into terms of a modified Harrod-Domar model (or more precisely, of the Feldman-Domar model) as done in the Appendix, and by putting it to work, we get the following results:

1. Other things being equal, an economy A with a higher share of its output coming from the capital goods sector (the Marxian "Department I") than the economy B will have a higher rate of economic growth, since it is able to make larger additions to its productive capacity over and above the current replacement needs. (In terms of original Marxian notation, the excess of $v_1 + s_1$ over c_2 is larger in A than in B.) Consequently, in order to increase its "warranted rate of growth," an economy would have to step up the relative share of the capital goods sector in its total output and in capital stock. Yet at this point the model would reveal its aspects which are more grim but also still more instructive.

2. To begin with, the model might be taken to mean that consumption cannot be lowered beyond what we could call the Von Neumann type of limit without reducing the number of workers manning the machines below the required level. (A more flexible interpretation would imply that a reduction in per capita consumption would negatively affect the efficiency of the given work force.) Yet this particular constraint on expansion could not have appeared to Marx as particularly severe, since he assumed that real wages lag behind the increase in the productivity of labor. Moreover, unlike his predecessors, he had no fear of diminishing returns in agriculture; and the notion of industrial development being obstructed by dependence on backward subsistence farming as the major source of food supply was undoubtedly still further from his thoughts.

3. But the model is much harder to assuage as far as the capacity side is concerned. The same logic that demands an increase in the share of sector 1 in the total income as a key to accelerated growth would make such increase dependent on the prior expansion of capital stock of this sector all the way up to the requisite level. (The full capacity utilization is assumed, of course.) Yet the rapidity of such an expansion would only partly depend on the extent of increase in savings that the economy would be able to undertake and to enforce. It would be decisively controlled by the relative share of the sector 1 in the total capital stock of the economy at the beginning of the process as well as by capital requirements per unit of new plant and its average gestation period. Given any halfway realistic numerical values for the ratio of total investment to the total capital stock and for the rates of speed at which stocks can be built (viz., run down by underreplacement), a marked shift in the sectoral composition of the economy must be a time-consuming process, even if pressed with utmost determination. As the illustrative example in our Appendix shows, it would (under the numerical coefficients adopted) take nearly two years to raise the relative size of the sector 1 to a level consistent with the doubled rate of growth. It is true that quite a few underlying assumptions—no foreign trade, full capacity utilization, no allowance for the possibility of conversion of some of the sector 2 plants for production of sector 1 goods—are unnecessarily harsh, and should be relaxed. Marx was fully aware of it;[11] but several assumptions of the opposite nature—gross investment plowed back in its entirety into the sector 1, with actual disinvestment in sector 2 as a corollary; average gestation period being equal to no more than one year and not getting any longer in spite of the sharpness of the switch[12]—are sufficiently breath-taking to provide a counterbalance, to say the least. Hence attempts to raise the rate of growth faster than capacity limitations permit are bound to be resisted. Marx referred to a similar situation when he noted that "transition from simple to expanded reproduction . . . will not always take place without difficulties" [12, p. 580].

[11] "We here take no account of foreign trade, by means of which a nation can change articles of luxury either into means of production or means of subsistence and *vice versa*" [10, p. 636, footnote]. Marx's admission that the number of workers per given plant can vary has already been mentioned; and the opportunities of direct conversion from "2" to "1" were of limited importance in an era when industries producing consumers' durables of the mechanical-gadget type did not yet exist.

[12] The lengthening of the average gestation period is a likely concomitant of such a switch for several reasons. The anticipated higher rate of growth would make it profitable, in many individual cases, to shift from the partial extention of the existing plant to more time-consuming full-scale construction of a new plant. This tendency would be particularly pronounced within the sector 1 industries whose share in the total output would have to be sharply increased; and the fact that the important basic-materials' subdivision of this sector has considerably higher gestation periods than the rest of the economy would tend to push up the average with added force.

And he came more directly to grips with the problem in a striking passage that sounded astonishingly like Keynes's much quoted statement about "pressure on the facilities for producing that type of capital [which] will cause its supply price to increase." "There is a check in reproduction and therefore in the flow of circulation. . . . The same phenomenon (and this as a rule precedes crises) can occur if the production of surplus capital takes place at a very rapid rate, and its retransformation into productive capital so increases the demand for all the elements of the latter that real production cannot keep pace, and consequently there is a rise in the prices of all commodities which enter into the formation of capital" [13, p. 371].[13]

The implications of all this for the issues we have been dealing with thus far are most significant:

1. The "displacement" problem acquires a new look, particularly in the early industrialization stages Marx was primarily confronted with. The opportunities opened up by new technologies and the external effects that went with them were striking—but the speed at which the young industrial economies could utilize these opportunities must have been narrowly circumscribed by limited capacities in the nascent capital goods industries and by the severe teething troubles these industries had to experience in process of breaking away from their artisan past. To put it in terms of the dichotomy suggested by Keynes and developed by Professor Lerner, while the marginal productivity of capital would be high and rising, the marginal efficiency of investment would be low and steeply declining. Hence the offsets against displacement may not have worked very effectively at first. By the same token, the situation was bound to change after the new sector 1 had expanded and reequipped itself. (Needless to say, we are dealing here with not implausible hypotheses and nothing else.) Besides, England, as the leading industrial country at the time when Marx wrote, could derive little benefit from importing major ingredients of her growing capital stock; indeed, her comparative advantages (geography-given and manmade) were particularly strong in the sector 1 area.

2. The abovesaid, if valid, fully bears on the broader issue of the rate of capital accumulation. In an economy with a low-keyed growth the problem of "pressure on the [capital-producing] facilities" is not

---

[13] E.g.: "A society where technical change and adaptation proceed slowly, where producers are reluctant to abandon traditional methods and to adopt new techniques, is necessarily one where the rate of capital accumulation is small. The converse of this proposition is also true: the rate at which a society can absorb and exploit new techniques is limited by its ability to accumulate capital" [6, p. 265]. The similarity between this passage and Marx's statements quoted in the footnote 10 (2) is evident. On the other hand, one could find in *Das Kapital* rudimentary elements of the "learning-by-doing" approach (cf. the brief quotation in the last paragraph of the preceding section, for instance) but their role in Marx's analysis should not be exaggerated.

likely to be serious even if the overall size of these facilities accounts for a relatively small fraction of the total stock. This is particularly true wherever a good part of investment activities occurs outside the modern sector, with labor being abundant, and nature not too forbidding; last but certainly not least, possibility of trading with the outside world on favorable terms would be helpful. But whenever an economy is experiencing a strong cyclical upswing, or whenever deliberate attempts are made to sharply lift the rate of growth from the hitherto prevailing level, the "ceiling" in sector 1 area are likely to make themselves felt sooner or later, more likely than not—in industries with particularly high capital-output ratios and long gestation periods. True, if the economy in question does have a substantial level of slack at the beginning of the process and if the inducement to expand is strong, the combined effects of multiplier and accelerator, involving a perceptible shift toward the sector 1 within the steadily growing investment volume, may carry the economy a long distance before the "ceilings" are hit; and either a better-than-average luck or a measure of realistic advance planning could go far toward eliminating them altogether. On the other hand, many booms die in their infancy and many leaps to higher steady growth rates fall short of target without ever getting close to ceilings of any sort. But whatever the case may be, the problem exists, and it has been widely recognized. The Marxian-type model, in our view, can be of great help in illuminating it.

3. The discussion of the preceding two paragraphs, overcompressed as it was, might convey a picture of greater rigidity than would be warranted by the facts of the case. To repeat, Marx explicitly noted important factors, making for greater flexibility: foreign trade, limited variability in degree of capacity utilization and in the service life of equipment, making use (even in process of partial replacement) of improvements in technology and, more particularly, of the fact that "when machinery is first introduced into an industry new methods of introducing it more cheaply follow blow upon blow" [10, p. 442]. Moreover, the services which the Volume II model performed for its Volume I counterpart were not unreciprocated. The whole "expanded reproduction" would be a monumental failure unless the growing plant capacity could find the labor to work with, and/or unless technological progress could keep reducing the labor requirements per unit of capital; and here, to be sure, the implicit accumulation model of Volume I would do its duty. Also the briefly sketched possibility of "dual economy" strategy could bring a measure of relief whenever the tension between investment opportunities and capacities would mount. Yet while all this could dull the knife edges, it would not transform them into slabs of butter.

IV

I have tried to demonstrate what the models contained in the two first volumes of *Das Kapital* could do for each other. Let me conclude by mentioning several things they could do for us:

1. The problem of technological displacement in leading industrial countries appears in a different light to us than it did to Marx. However, this difference can be to a considerable extent explained in terms of his own analysis. It is still true that "not enough means of production are produced to permit the employment of the entire able-bodied population under the most productive conditions" [13, p. 302]. Yet the importance of the capacity as constraint on employment, even under less than "most productive conditions" has dramatically declined since Marx wrote. The capital stock in the developed countries has grown in size much faster than the labor force, and the sector 1 of the economy is now in a much better position to create adequate "offsets" against displacement tendencies. Hence the "industrial reserve army" can be absorbed into the system, provided that the aggregate demand is large enough. Nevertheless, noncompensated displacement might still occur if (a) technological progress is sharply labor-saving; (b) a considerable portion of new equipment is coming from plants that had been producing the old equipment and can use essentially the same sources of energy; to that extent the buoyant force of the "new industry effect" to which Marx (and Schumpeter) attributed such a great importance would be lost; (c) the capital stock of the economy was growing rather slowly over a long period of time; and (d) a substantial part of the sector 1 potential is preempted by military demands.

2. All this, if true, can establish only a possibility of technological unemployment. (Point (b) is certainly true only in part; but this might mean that the peak of unemployment would merely be shifted to the period when "new industries" had already met the bulk of reequipment demand.) To rule it out would be rash. Yet after all is said and done, it is understandable that the attention of some of the leading economists of our days moved back one link along the causation chain forged by Marx. The most striking instance of it can be found in Professor Robinson's recent writings which stress the role of labor scarcity caused by rapid accumulation as a powerful force pushing for technological progress. Space precludes elaboration of this point. I can do little more than record my agreement with Professor Robinson in her insistence that the emphasis on this relationship had been one of Marx's signal contributions, and note that here too the compliment is being returned, with Marx rescuing the original Keynesianism from

some of its overstatements. The notion of investment of today digging the grave of the investment of tomorrow loses some of its drama if it can be shown how this very investment by pressing against a less rapidly increasing labor force, can propel the economy toward a new production function and thus generate investment opportunities in excess of what a mere sliding along the old one could provide. (Marx, as was repeatedly pointed out, made no allowance for the possibility of such "sliding along," while strongly emphasizing the none-too-peaceful coexistence of technologies of various vintages. But the neo-Keynesians have gone a long way toward him also in this respect when they stressed that the notion of a movement along a given production curve is of highly limited relevance for tracing the path of investment over time as distinct from describing individual investment choices at a given point of time, and when they effectively challenged the notion of the aggregate production function for the whole economy.) This, to be sure, need not imply that investment opportunities thus created will be necessarily sufficient to lift the economy in question all the way toward full employment level and to keep it there, or that all technological progress is induced by labor scarcity; Marx certainly did not subscribe to such a monistic view either. But also those theorists who see the operative link between investment and technological progress in "learning by doing" rather than in labor scarcity (with others like Professor Robinson emphasizing both) frequently describe the overall interdependence between the two phenomena in a very "Marxian" manner. I refer primarily to the recent writings of Professor Kaldor.[14]

3. A brief postscript on the Volume II model seems in order. To begin with, a "ceiling" of the type described in the preceding section is a short-run concept, and it cannot therefore be expected to perform all the services of the "law of the falling rate of profit." Yet its role, while more modest and less doom-laden, is far from insignificant. By bringing into focus the basic fact that investment not only adds to the capacity but also presupposes a capacity of certain size and structure, the two-sectors construct injects an element of realism in our notions about the plausible speeds (or rather about plausible rates of change in speed) of accumulation processes. More particularly, it helps us to understand the "stop-go" pattern these processes are likely to produce when entrepreneurs are Schumpeterians or when central planners are Stalinists. (The connection between overambitious plans and quasi-cyclical fluctuations in the rate of growth of the Soviet-type economies was sugges-

[14] It goes without saying that analogy between the "ceilings" in market economies and in the Soviet-type economies should not be pressed too far. Differences in impact and in the mode of operation are no less important than similarities. (I have briefly discussed both in my paper, "Development Strategy and Planning: The Soviet Experience," to be published in the forthcoming Universities-NBER volume, *National Economic Planning*.)

tively discussed by the Czechoslovak economist Josef Goldmann who also coined the *mot:* "Big leaps belong in the gym." I owe the reference to this phrase to Professor Holesovsky).[14] It is quite true that developed economies are likely to be more successful than underdeveloped in overcoming some of the rigidities of the model. But here the situation is analytically not unlike the case of technological unemployment, or in fact "even more so." Relaxation possibilities which have been duly listed before work much better for small and shortlived changes in the rate of growth than for discontinuous and enduring ones; and they are obviously not unlimited in any case.

Much of what has been said here represents, in all likelihood, a minority view, with the inevitable overcondensation making it, I suspect, sound rather dogmatic in spots. It would be incongruous to end on a note of consensus even if the term had not been so tarnished nowadays. And yet a basis for a limited agreement does seem to exist. It is a fact that Western economists of widely varying persuasions are now wrestling with problems posed by Marx instead of gingerly relegating them to the realm of "data." It is incontrovertible that in the Soviet Union and in Eastern Europe the revival of serious macroeconomic thinking along Marxian lines marks one of the major paths of return to intellectual integrity and sanity. This should make it easier for all of us, East and West, to take full measure of the man who refused to rig his assumptions to suit his purposes and had the giant's vision as well as the giant's heart to see the system he detested in its open-ended complexity. It seems therefore not too much to expect that many who are less inclined to accept important elements of Marxian analysis than I am will join in honoring the creator of *Das Kapital* on the eve of its centenary.

APPENDIX

Notation:

K—capital stock
Y—national income
I—net investment
R—replacement
GNP—$Y+R$
"1"—capital goods' sector
"2"—consumer goods' sector
r—annual rate of growth
s—I/Y
V_y—K/Y
V_{GNP}—K/GNP

Steady Growth:

$$r = 5\%; \quad s = 15\%; \quad v_y = 3; \quad v_{GNP} = 2.5$$

Year 1	K	Y	I	R	GNP
"1"	87.5	29.2	4.4	5.8	35
"2"	212.5	70.8	10.6	14.2	85
Total	300.0	100.0	15.0	20.0	120

Year 2					
"1"	91.9	30.6	4.6	6.1	36.7
"2"	223.1	74.4	11.1	14.9	89.3
Total	315.0	105.0	15.0	21.0	126.0

Accelerated Growth:

r is to increase from 5 percent to 10 percent, with V_y and V_{GNP} assumed constant. Consequently s must increase from 15 percent to 30 percent, in keeping with the familiar Harrod-Domar formula and the share of "1" in K, Y and GNP must go up accordingly. It is further assumed that, in order to carry out this increase at maximum speed, the whole $I+R$ is channeled toward "1" for the duration of the adjustment.

Year 1	K	Y	I	R	GNP
			As above		
Year 2					
"1"	116.7	38.9	9.6	7.8	46.7
"2"	198.3	66.1	16.2	13.2	79.3
Total	315.0	105.0	25.8	21.0	126.0
Year 3					
"1"	155.6	51.9	18.0	10.3	62.2
"2"	185.2	61.7	21.5	12.4	74.1
Total	340.8	113.6	39.5	22.7	136.3

$$s = \frac{39.5}{113.6} = 34.8\%; \quad r = \frac{34.8}{3} = 11.6\%$$

Conclusion: under numerical assumptions of our example, it takes slightly less than two years to adjust the size and structure of the capital stock of the economy to the desired rate of growth, after which I'/I can again become equal to K'/K.

References

1. M. Bronfenbrenner, *"Das Kapital* for the Modern Man," *Sci. and Soc.,* Fall, 1965.
2. ————, "The Marxian Macro-Economic Model: Extension from Two Departments," *Kyklos,* June, 1966.
3. M. Dobb, *Political Economy and Capitalism* (New York, 1945).
4. E. D. Domar, "A Soviet Model of Growth" in *Essays in the Theory of Economic Growth* (New York, 1957).
5. N. Georgescu-Roegen, "Mathematical Proofs of the Breakdown of Capitalism," *Econometrica,* Apr., 1960.
6. N. Kaldor, "A Model of Economic Growth," in *Essays on Economic Stability and Growth* (London, 1960).
7. O. Lange, "Marxian Economics and Modern Economic Theory," *Rev. of Econ. Studies,* June, 1935.
8. ————, *Introduction to Econometrics* (2nd ed., Oxford-London-New York, 1962), Chap. III, and Appendix II.
9. A. Lowe, "Structural Analysis of Real Capital Formation" in Universities-NBER, *Capital Formation and Economic Growth* (Princeton, 1955).
10. S. H. Mage, "The Law of the Falling Tendency of the Rate of Profit in the Marxian System" (Ph.D. thesis, Columbia Univ., 1963)
11. K. Marx, *Capital,* Vol. I (Chicago, 1906).
12. ————, ————, Vol. II (Chicago, 1909).
13. ————, ————, Vol. III (Chicago, 1909).
14. ————, *Theorien Uber den Mehrwert,* Vol. II/2 (Stuttgart, 1905).
15. ————, ————, Vol. III (Stuttgart, 1910)
16. ————, *Theories of Surplus Value* (New York, 1952).
17. J. Robinson, *An Essay on Marxian Economics* (London, 1947).
18. R. Rosdolsky, "Zur neueren Kritik des Marxschen Gesetzes der fallenden Profitrate," *Kyklos,* 1956, No. 2.
19. P. A. Samuelson, "Wages and Interest: Marxian Economic Models," *A.E.R.,* Dec., 1957.
20. A. Smith, *The Wealth of Nations* (Modern Library ed., New York, 1929).
21. P. M. Sweezy, *The Theory of Capitalist Development* (New York, 1956).

MARXIAN ECONOMICS AS ECONOMICS

By Paul A. Samuelson

Massachusetts Institute of Technology

Genius or Crank?

This coming year Marx's *Das Kapital* celebrates its hundredth anniversary. At such a birthday party, only the Good Fairies should be invited. Those who cannot find anything at all nice to say should decline the invitation. On the other hand a great scholar deserves the compliment of being judged seriously; and truth does have its claims, on holidays as well as working days.

The "contradictions of capitalism," which Karl Marx saw everywhere, are as nothing compared to the contradictions of Marx himself. Marx was a gentle father and husband; he was also a prickly, brusque, egotistical boor. (Even Engels, his ever faithful friend, found it too much when Marx greeted the news of the death of Engels' working-class mistress with the callous response that now more work could be got done.) Although Marx was a learned man, he shows all the signs of a self-taught amateur: overelaboration of trivial points, errors in logic and inference, and a megalomaniac's belief in the superiority of his own innovations. He introduced into scholarly literature manners not seen since the polemics of the renaissance. Too bad Marx could not have done systematic graduate work at Harvard under John Stuart Mill, and then been given a good chair at Columbia!

Evaluations of Marx show the same pattern of contradictions. Professor Bronfenbrenner, my colleague on today's platform, deems Karl Marx "the greatest social scientist of all times." Keynes consistently refers to the "turbid rubbish of the Red book stores" and dismisses the book we commemorate today as a "bible, above and beyond criticism, an obsolete textbook which I know to be not only scientifically erroneous but without interest or application for the modern world." This attitude Joan Robinson regards as rather a pity, saying: "Keynes could never make head or tail of Marx. . . . But starting from Marx would have saved him a lot of trouble [as it did Kalecki]." In my Presidential Address, I find Marx referred to as "from the viewpoint of pure economic theory, . . . a minor post-Ricardian . . . a not-uninteresting precursor of Leontief's input-output."

There you have a spread of opinion—from the greatest social scientist to purveyor of rubbish. To ask what view is right is like asking whether the box in an optical illusion is inside-out or outside-in. There

is no test-of-truth by which bets could be settled about the correctness of one view rather than another. Let me, therefore, turn my microscope onto aspects of Marxian economics that can be fruitfully discussed. But not before mentioning a reason why, beyond his scientific merits, we find a man like Karl Marx worth discussing.

For better or worse, Marx is an important figure in the history of ideas. And much is known about him—his fugitive letters, juvenile manuscripts, I dare say even his laundry lists. When a sizable audience knows much about a man—whether he be Dr. Samuel Johnson, Sherlock Holmes, or Karl Marx—the facts about him become subject to the law of increasing marginal utility. Frederick the Great's flute compositions would not sell as well under any other name. Most of Samuel Johnson's ideas were really pedestrian; but after we have pored much over his countenance, his face becomes like that of one of the family and each wrinkle takes on an interest all its own. Many a newly published fragment by Marx would be of no interest at all if known to be the work of some 1844 John Doe; the whole becomes greater than the sum of its parts—not because the Bronfenbrenner quotation from Veblen about the organic coherence of the Marxian system is really true, and not even because each fragment contributes something to the grand symphony of his thought, but merely because of an antiquarian interest that becomes like a detective-story game. Camp is a new word for an old—and, I may add, defensible—preoccupation.

But back to my microscope.

Tableaux of Expanded Reproduction

First, we can make a deposition—as the lawyers say—that Marx did, in his posthumous Volume II, innovate two-sector models of reproduction and growth. These are useful anticipations of work done in our day by Harrod, Domar, Leontief, Solow, Robinson, Uzawa, Pasinetti, Kaldor, Findlay, and many others. I do not honestly think that modern developments were much influenced, directly or indirectly, by Marxian writings; instead they grew naturally out of a marriage of the Clark-Bickerdike accelerator and the Keynes multiplier, and out of earlier works by Von Neumann and Frank Ramsey that show no Marxian influence. But still we all might well have benefited earlier from study of the Marx tableaux.

Second, there is a point made by Leontief himself. Many of these same Marxian models stressed the role of fixed capital in a way that the Austrian School generally did not. Because Böhm-Bawerk tied himself to simple arithmetic examples, his *Positive Theory of Capital* is almost always expressed in terms of circulating-capital models of goods-in-process. For Böhm, labor alone produces goods in the earliest stage

of production—say wheat. Then labor and wheat produce dough. Then labor and dough produce bread. There is no explicit need for durable capital goods in this "hierarchical" structure of Austrian production. (In terms of Leontief input-output, the a_{ij} matrix is not only "triangular," permitting classification of goods into "earlier" and "later"; also, each good depends only on one earlier good, with all a's zero except, $a_{i-1, i}$.)

Marx on the other hand considered bread as being produced by labor and ovens; and ovens as being produced by labor and ovens. In Leontief's 1937 A.E.A. address on Marx, this is rightfully hailed as an important innovation. As Adolph Lowe and the late Frank Burchardt have stressed, the Leontief flow of circular interdependence is more Marx-like than Austrian.

Leontief refers to the "rather paradoxical situation. The dean of the bourgeois economists [Böhm] insisted on theoretical reduction of all capital goods to pure labor; he was opposed by the formidable proponent of the labor theory of value [Marx] in the role of a defender of the independent, primary function of fixed capital."

Leontief is calling attention here to a deeper paradox than that involved in the spectacle of a French Marxist advising the Indian government that labor is a redundantly free factor and capital alone is scarce—all having to be couched in terms of the concepts of the labor theory of value, a Yoga-like feat worthy of Hercules. Leontief goes on to claim superiority for the Marx model to handle the problem of high-wage-induced-substitution-of-machinery-for-labor. But is Leontief right in this contrast? In 1937 Leontief had not yet had the chance to remember the 1949 Nonsubstitution Theorem for the Leontief system. According to it, if the rate of interest or profit stays the same, that money wage increase which raises all prices proportionately in the Austrian wheat-dough-bread system will also raise all prices proportionately in the Leontief-Marx nontriangular system. Long-run substitution comes in either system only if the equilibrium interest rate changes.

Marx's model of expanded reproduction is perhaps the first example of those golden-age paths of compound interest which Cassel, D. H. Robertson, Von Neumann, Harrod, Domar, and all the rest have made so fashionable in modern economics. Before leaving it, let us note that it could lend substance to Marx's jest: "I am not a Marxian." Using it, he could say, "I'm not a post-Marxian of the Luxemburg underconsumptionist type." With historians Marx is able to have his cake and eat it too. On the one hand, he is the Ricardian critic of Malthusian underconsumptionist notions held by contemporary socialists like Rodbertus; on the other hand, he is hailed as a precursor of Keynes (and Major Douglas, Gesell, Hobson, Foster, etc.). Can a scholar have it

both ways? In this respect, how can you be a precursor of Keynes without being a postcursor of Malthus? Perhaps being confused helps.

In any case, the compound interest rates of growth of the reproduction tableaux can provide the way out of some dilemmas of ultimate underconsumption that bothered Rosa Luxemburg and later Marxists. (See Paul Sweezy's valuable *Theory of Capitalist Development,* particularly Chap. X and its Appendix.) If accumulation of profits can just suffice to keep all magnitudes growing in balance with smoothly growing labor supply for a few periods, compound interest says it can continue to do so forever. Many of the demonstrations to the contrary foundered on linear rather than exponential examples. (Yet, remember that saving and accelerator coefficients must be right in the beginning if the "warranted" growth rate is to just match the "natural" growth rate of labor so that the same behavior relations can be assumed to hold indefinitely; unless, as in bourgeois economics, there is a mechanism that causes such saving-accelerator coefficients to adjust to the requirements of equilibrium, it is an improbable razor's edge case in which the Marxian tableaux can step off in equilibrium.)

The Labor Theory of Value

As every encyclopedia reader knows, Marx believed in the labor theory of value. One might expect me at this stage of the birthday party to examine its demerits. But the many economists speaking on these platforms of the American Economic Association have examined its demerits far beyond my poor powers to add or detract. Let me therefore be dogmatically terse.

Proposition 1. Adam Smith held a labor theory of value for about as long as it takes a grown man to turn two pages of his book. David Ricardo never shook himself free of this incubus, but no reader of Sraffa's edition can fail to be persuaded that only some of the simplified numerical examples in the Ricardian system need have any reliance on such a theory.

Proposition 2. From the standpoint of science, the labor theory of value breaks down even before complications of capital enter into the model. With land scarce and different goods varying in their labor-land intensity, already goods will exchange at relative prices that are not proportional to socially-necessary labor content. Ricardo nodded and thought that by going out to the external margin of no-rent land, he could "get rid of the complication" of land costing. Why should we, or the Soviet planners, nod with him? (This point is obvious and appears in the first pages of the new edition of my *Economics;* yet when I searched the literature of the labor theory of value for it years ago, I could turn up only one reference to Lionel Robbins.)

Proposition 3. If Marx had intended to use the labor theory of

value to lay bare the laws of motion of capitalism and if he had been barking up the right tree, then the inadequacies of the labor theory of value as exposited in Volume I of *Capital* would not really have mattered.

Let me explain what I mean. Most of Volume I would stand up if Marx stipulated, purely for expository simplicity, that the organic composition of capital (or as we would say, labor's fractional·share of value added) were the same in all industries. By fiat the contradiction between equal rates of surplus value and equal rates of profit would disappear. (And make no mistake about it, Böhm-Bawerk is perfectly right in insisting that Volume III of *Capital* never does make good the promise to reconcile the fabricated contradictions. When Paul Sweezy says that Rudolf Hilferding, in refuting Böhm's specific critiques of Marx, "gives a good account of himself and shows that even at the age of twenty-five he could stand up and trade punches with so experienced and inveterate a polemicist as Böhm-Bawerk," I have to pinch myself to remember that relative prices of goods do really change as demand changes even when their socially-necessary labor contents do not change—which is all the dispute is really about.)

In 1865, when Marx was at the height of his powers and had to boil down the message of his masterwork for a workers' audience, he introduced into the pamphlet, *Value, Price and Profit,* the simplifying notion that prices are proportionate to labor values—saying "apart from the effect of monopolies and some other modifications I now pass over." I suggest that much ink and blood would have been spared if he had done likewise in *Capital.* When a modern theorist assumes equal factor intensities in a two-sector Ramsey-Solow model, he does not defend the oversimplification: he is content to know that anything interesting turned up in it is likely to be of relevance for a more complicated model.

In summary, if labor-theory-of-value reasoning, as applied to an impeccable model of equal factor intensities, turned up new light on exploitation in an existing system or if it turned up new light on the laws of development of capitalism, it would be an invaluable tool even though not defensible as a general theory of markets.

If, and if. Let us see whether Marx was at all barking up the right tree.

Laws of Motion of Capitalism?

The usual claim for superiority of Marx's system is not that he beats the vulgar economists at their own game of describing equilibrium pricing, but that their game is not worth the playing: whereas Wicksell, Walras, and Chamberlin give a good enough description of

the economic system as it is, we must turn to the Marxian system for insight into the laws of development of the capitalistic system. Its inferior statics can be forgiven considering its much superior dynamics. Such a claim, if it can be sustained, is indeed a weighty one.

Let us review the authorities. Leontief, in that same 1937 address, makes heavy weather of finding much to praise in Marx besides his anticipations of input-output. But Leontief is able to say:

However important these technical contributions to the progress of economic theory, in the present-day appraisal of Marxian achievements they are overshadowed by his brilliant analysis of the long-run tendencies of the capitalistic system. The record is indeed impressive: increasing concentration of wealth, rapid elimination of small and medium sized enterprise, progressive limitation of competition, incessant technological progress accompanied by the ever growing importance of fixed capital, and, last but not least, the undiminishing amplitude of recurrent business cycles—an unsurpassed series of prognostications fulfilled, against which modern economic theory with all its refinements has little to show indeed.
Neither his analytical accomplishments nor the purported methodological superiority can explain the Marxian record of correct prognostications. His strength lies in realistic, empirical knowledge of the capitalist system. (*A.E.R.*, Mar. sup., 1938, pp. 5, 8.)

Here Leontief is referring to the then recent work by Oskar Lange, whose death we have so recently mourned. The years 1934 to 1944 constituted Lange's wonder decade, during which he turned out brilliant articles in capital theory, welfare economics, Keynesian model building, and much else. In the 1935 *Review of Economic Studies,* Lange compares the merits of Marxian and modern economics and finds Marxian economics superior in specifying the institutional data out of which can be formed a theory of capitalistic development. Despite its outdated concepts, Marxian economics is believed by Lange to be able to explain what bourgeois economics has utterly failed to explain: "the fundamental tendencies of the development of the Capitalistic system—the constant increase of scale of production leading to the present monopolistic (or rather oligopolistic) Capitalism; the substitution of . . . 'planning' for *laissez faire;* . . . free trade to protectionism; . . . imperialist rivalry among the principal capitalist powers; increase of economic instability leading to rebellion (Socialism or Fascism)."

Here Lange is proceeding from the 1933 *Kyoto Economic Review* article by Kei Shibata, which asserted that Marxian political economy "sets forth theories which . . . enunciate systematically the organisation of present-day capitalistic society and the laws governing its development." As I understand him, Lange is agreeing with the dynamic superiority of Marxian economics and seeking its source; but, unlike Shibata, he does not concede its superiority to explain the then current economy. For Lange points out current "problems before which Marxian economics is quite powerless. What can it say about monopo-

ly prices? . . . monetary and credit theory? . . . incidence of a tax, or of technical innovation on wages?"

You will notice that Leontief credits Marx with great prophetic powers but is noncommital as to whether Marx's economic theories helped him to arrive at these (possibly merely lucky) guesses. Lange attempts to make stronger claims for Marxian theories. He says they deduce that "the fundamental change occurs in production and that the 'necessity' of such a change can be deduced only under the institutional set-up specific to Capitalism. Thus a 'law of development' of the Capitalist system is established . . . not a mechanical extrapolation of a purely empirical trend. . . . "

So much for the claims. But is it so? Let us be honest children and ask whether the Emperor is really wearing clothes, and whether those clothes really do follow some grand theoretical pattern.

Specifically, was Marx right as a prophet of the future of Victorian capitalism? The immiserization of the working class, which he thought to deduce from the labor theory of value and his innovational concept of surplus value, simply never took place. As a prophet Marx was collosally unlucky and his system collosally useless when it comes to this key matter. This is not to deny Joan Robinson's view that such a prophecy had a certain propagandistic value. She says, "This error, like Jesus' belief that the world was shortly coming to an end, is so central to the whole doctrine that it is hard to see how it could have been put afloat without it. . . . 'You have nothing to lose but the prospect of a suburban home and a motor car' would not have been much of a slogan [for the Communist Manifesto]." With friends like this, who has need for an enemy?

Let's now move on to the growing monopolization under capitalism. For thirty years Marx seemed to have been right in this prophecy, even though for the next seventy years he does not seem to be borne out by the most careful of researches on industrial concentration. But suppose he (and numerous non-Marxian socialists) had been right in this view. Would such an extrapolation be deducible in any way from the surplus value ratios, $S/(V + C)$, of any of the volumes of *Das Kapital?* No one has yet shown how, and I have to agree with the recent book of Paul Sweezy and Paul Barran which seeks to identify as an important explanation of the stagnation of Marxian social science the fact that "the Marxian analysis of capitalism still rests in the final analysis on the assumption of a competitive economy" (*Monopoly Capital*, 1966, p. 4).

Since time is short let us rush on to consider whether it is an inevitable law of capitalist development that the business cycle should be getting worse and worse. Shibata and Lange, writing in the 1930's, might

be forgiven for thinking so, just as writers in 1929 can have been expected to celebrate the demise of economic fluctuations. Who can blame someone for not having predicted in 1867 the successful development of the Mixed Economy, in view of the fact that so astute a philosopher as Joseph Schumpeter managed to miss foreseeing it as late as 1947? I throw no stone at Marx, because I have never believed in the big-picture theories of anyone—Toynbee, Spengler, Schumpeter, Veblen, Marx, or even Rostow and Galbraith. But those who have been bewitched by a belief in the timetable of history, as deduced by theoretical laws of motion of capitalism, should taste the bitter bread of disillusionment.

Had Lange been writing in 1937, after Keynes, he might have added to the 1935 sentence "Marxian economics would be a poor basis for running a central bank or anticipating the effects of a change in the rate of discount" the sentence, "and it would be a poor basis for understanding the role of fiscal policy in maintaining high employment." What admissions! This is equivalent to saying, "Marxian economics is powerless to explain the 1937-67 developments of European and American economies."

The cash value of a doctrine is in its vulgarization. To understand the pragmatic content of Marshall, you must read Fairchild, Furniss, and Buck. To prove the Marxian pudding, only read the Soviet textbooks dealing with American and Western economic systems. Aesthetics aside, their predictive powers have been unbelievably erratic and perhaps only to be understood in terms of the dictum: Marxism has been the opiate of the Marxians.

But this is a birthday party and I approach the boundaries of good taste. Let me conclude by wishing that, like Tom Sawyer attending his own funeral, Karl Marx could be present at his own centennial. When "the Moor" rose to speak, how we would all pay for our presumptuousness!

MARXIAN INFLUENCES IN "BOURGEOIS" ECONOMICS

By Martin Bronfenbrenner
Carnegie Institute of Technology

"Why on earth should a man, because he is a Marxist, be a drivelling idiot?"
—Boris Pasternak, *Doctor Zhivago.*

I

Das Kapital's centenary finds Karl Marx still a controversial figure, wherever he is neither a plaster saint nor a four-letter word. Nowhere has he been easily forgotten, and *Das Kapital* is still the most influential unread book in existence.

Precisely because Marx is a controversial figure, let me state in advance my personal bias concerning him. This bias is expressed most readily by analogy. Suppose one asked a sample of Unitarian ministers their choices as the greatest religious philosopher of recorded history. Most would vote, I am sure, for Jesus, Buddha, or Moses, and yet a Unitarian is neither a Christian nor a Buddhist nor a Jew. In the same way, were I personally asked to name the greatest social scientist of all time—not necessarily the greatest economist—I should name Karl Marx, but without considering myself a Marxist or being considered one by my exclusivist Marxian friends. There are too many "bourgeois" elements in my thinking, however great my admiration for the Marxian theoretical structure, and I remain a muddled eclectic. (F.B.I. and Birch Society please note.)

II

My assignment, to consider bourgeois economics' debt to Marx, recalls this Association's last full-dress "Marxism" session, in 1937. At that time Wassily Leontief and the late Leo Rogin agreed that both contemporary orthodoxy and the early New Deal owed considerably less to Marx than many anti-intellectual and anti-Roosevelt extremists supposed at the time.[1] With Rogin's "policy" verdict I shall not disagree, but Leontief's "theory" verdict might well be revised upward in view of our own advances since 1937. Which leads me to wonder, may

---

[1] Wassily Leontief, "The Significance of Marxian Economics for Present-Day Economic Theory," and Leo Rogin, "The Significance of Marxian Economics for Current Trends of Government Policy," John Ise and Joseph J. Spengler, discussants (*A.E.R.*, Mar. sup., 1938). Stronger views than these are quite commonplace. I cite at random Raymond Aron, "The Impact of Marxism," in Milorad M. Drachkovitch (ed.), *Marxism in the Modern World* (Stanford Univ. Press, 1965), p. 15, as a strong statement with which I propose to take issue: "To profit from the progress . . . in economic thinking since Ricardo and Marx, one would have had to dispense with the conceptual apparatus of *Das Kapital.*"

not further upward revision of my own estimate be required, if it is exhumed in 1997?

I propose to modify my assignment in two ways. First, by limiting myself to the debt owed the Marxian system by the so-called "conventional" wisdom of our profession and omitting any consideration of the relation between Marx and the dissidents who preceded and followed him. Second, by considering not only the debt we owe Marx in fact, which now appears to me greater than it did to Leontief a generation ago, let alone to Böhm-Bawerk and Thorstein Veblen a generation before Leontief, but also the debt we should have owed him from the outset had his ideas been more felicitously phrased and our predecessors more willing to listen to them.

These modifications may require defense. The problem immediately arises, in the first place, of separating out Marxism specifically from the wave of economic heterodoxy, socialist and non-socialist, which has served as antithesis to the great theses of first the classical and then the neoclassical schools. This problem I lack scholarship to solve, and I should prefer to interpret my function in such wise as to enable me to dodge it instead.

Marx was, like Keynes, primarily a synthesizer, at least in his economics. There are few if any elements of his system which cannot be found in embryo in one or another predecessor. (The English "Ricardian Socialists" come immediately to mind.) There are yet fewer elements not paralleled by one or another reformist or socialist contemporary or near-contemporary. Marx's genius lay, like Keynes's, in synthesis, in combining bits and pieces from one and another system into a whole greater than the sum of its parts. It is this aspect of Marxism, in particular, that seems to have gone unappreciated by Marx's earlier bourgeois critics, who tend to hack away at trees without disturbing the forest.[2] It is worth pausing to observe those modern and ultramodern constructs which went unrecognized for two generations or more. At the same time, who can say whether some quasi-Marxian influence in conventional economics came directly from Marx or from any of half-a-dozen sources independent of Marx, including the *Zeitgeist?*

III

Before getting down to the substance of this paper, besides redefining my assignment for my own purposes, I should like to help

[2] However, consider Veblen, "The Socialist Economics of Karl Marx and His Followers," *Q.J.E.*, Aug., 1906, reprinted in *The Place of Science in Modern Civilization* (New York: Huebsch, 1919), p. 410 f. "Except as a whole . . . , the Marxian system is not only not tenable, it is not even intelligible. . . . No member of the system, no single article of doctrine, is fairly to be understood, criticised, or defended except as an articulate member of the whole and in the light of the preconceptions and postulates which afford the point of departure and the controlling norm of the whole."

lay an extraordinarily durable ghost. This ghost is the perverse influence some people suspect that Marx exercised on the subsequent development of theoretical economics. It is the belief that the subjective, marginal, or utility revolution in value and price theory was prompted ideologically, to escape from the consequence of the labor theory of value as developed particularly by Marx.[3] This thesis is not proven; in fact, the weight of evidence seems to be against it.

Offhand, the case looks suspicious in both time and place. Volume I of *Das Kapital* appears in 1867. The year 1870 is the accepted date for the Jevons-Menger-Walras utility revolution, which carried the field where earlier efforts along identical lines had apparently been dismissed out of hand. Moreover, the labor theory failed more rapidly in the German-speaking countries, where Marxism was strong, than in the French- and English-speaking ones, where it was weaker. (The successor to the classical labor theory was not uniformly marginal utility. Most notably in Germany, the residuary legatee was some form of historicism, but that is another issue.)

The main evidence against the Marxophobe thesis regarding the development of utility theory seems to be that *Das Kapital* itself succeeded so slowly, except in primitive, precapitalist Russia. "In Western Europe, Engels had to write virtually all the reviews . . . under his own name and various pseudonyms."[4] By the time the first volume became reasonably well known, in the 1880's, the utility revolution was independently in full swing, and marginal productivity was peering marginally over the marginal horizon.

[3] Two examples will suffice, one from an eminent economist and the other from an outsider with iconoclastic ambitions: "It is worth recalling that parts of [the marginal productivity theory] were, to some extent, originally developed to provide a rebuttal to Marx's theory of exploitation." Tibor Scitovsky, "Some Theories of Income Distribution," in *The Behavior of Income Shares* (Princeton Univ. Press for National Bureau of Economic Research, 1964), p. 22; "[E]ver since Marx used Ricardo to expound his famous labor or surplus theory of value (a thunderous moral statement) traditional economics has noticeably lost interest in what was formerly the central problem in economic theory, the problem of value." David Bazelon, *The Paper Economy* (Random House, 1963), p. 15.

[4] Bertram D. Wolfe, *Marxism* (Free Press, 1965), p.x. Engels wrote at least nine separate reviews of the first volume of *Das Kapital* (*ibid.*, n. 3). The British Marxist historian E. J. Hobsbawm writes of British reaction: "Between 1850 and 1880 it would have been hard to find a British-born citizen who called himself a socialist in [the modern] sense, let alone a Marxist. The task of disproving Marx was therefore neither urgent nor of great practical importance. . . . [Although the earliest non-Marxist 'expert' on Marx wrote in 1879], "I doubt whether anything even approximating to a usable non-socialist summary of the main tenets of Marxism . . . exists before Kirkup's *History of Socialism*" (1900). "Dr. Marx and the Victorian Critics," in *Labouring Man: Studies in the History of Labour* (London: Weidenfeld and Nicolson, 1964), p. 240 f. Professor Herbert G. Gutman has supplied me with parallel American information as well. For example, the Chicago *Tribune* warned against socialism in December, 1873, for the reason (among others) that Marx had rejected classical economics! Three years later, a New York labor paper (*The Socialist*) began summarizing the first volume of *Das Kapital* as a weekly serial. Gutman, "Failure of the Movement by the Unemployed for Public Works in 1873," *Polit. Sci. Q.,* June, 1965, p. 272.

As for the German anticlassical revolution, it was well under way by 1867; in fact, one of Marx's favorite vulgar-economist whipping-boys was none other than Wilhelm Roscher, best known to doctrinal history as a principal founder of the "older" historical school.

There is a subordinate point of similar import. Marx was no shrinking violet regarding his own importance in the history of thought, but both he himself and his followers have pointed to Ricardo, his predecessor, as having frightened the bourgeoisie with the implications of the labor theory of value and induced its weakening and abandonment by the vulgar economists. *Theorien über den Mehrwert* is of course the text here, and subsequent Marxian and neo-Marxian accounts of doctrinal history take a similar line.[5]

IV

In considering what students should know about Marxian macroeconomics (in un-Marxian isolation from the remainder of Marx's social philosophy) I have found two expository devices both effective and time-saving: (1) formulating Marxian statics as a simple Lausanne school general-equilibrium system, and (2) formulating Marxian dynamics in a "dilemma" diagram, by which no profit rate could remain, as technology progressed with a laborsaving bias, simultaneously high enough to avoid liquidity crises and low enough to avoid overproduction at (or below) any predetermined unemployment percentage. These devices or "Marx-like models" having appeared in print elsewhere,[6] I relegate them, in modified form, to an Appendix. Here, with occasional references to specific equations of this Appendix, I

[5] For example, Rogin argues, discussing Jevons (*The Meaning and Validity of Economic Theory*, Harper, 1956, p. 468 f.): "Ricardo's theory of the inverse relation [between wages and profits] . . . was developed incidentally to his preoccupation with the trend of profits as the criterion of economic progress. . . . But with the shift from the social and political conflict between landlords and the industrial bourgeoisie to the one between labor and capital, Ricardo's theory [N.B. Not Marx's theory] served to feed theoretical fuel to the flames of the latter conflict." A more elementary Marxist writer, John Eaton (*Political Economy* [New York: Int. Pub., 1966], p. 27) is more explicit: "Marxist economic theory was built upon the scientific foundations laid by . . . Adam Smith and Ricardo, for whom the labor theory of value was the foundation of economic science . . . [T]he labor theory of value enabled Marx to show the nature of capitalist exploitation and that capital itself was doomed to extinction. The defense of capitalism called, therefore, for an attack upon the labor theory of value. Bourgeois theory was quick to sense this, and from about 1830 [N.B. Not 1867] has been in quest of an economic theory that rejected the labor theory of value."

[6] M. Bronfenbrenner, "*Das Kapital* for the Modern Man," *Sci. and Soc.*, Autumn, 1965, "Classical and Marxian Macro-Economics in Separate Nutshells," in *Essays in Honour of Marco Fanno* (Padua: Cedam, 1966), pp. 140-50, and "The Marxian Macro-Economic Model: Extension from Two Departments," *Kyklos*, June, 1966. My debt to Lawrence R. Klein,"Theories of Effective Demand and Employment," *J.P.E.*, Apr., 1947, will be obvious. Candor also requires the admission that a proposal for translation of the *Science and Society* essay into German has been rejected (in East Germany) because of alleged distortions of Marxian doctrine. The modifications in the present Appendix reflect primarily criticisms received at a Johns Hopkins University seminar in May, 1966.

limit myself to a catalogue of some nine "modern" elements of *Das Kapital,* not all of which I find personally congenial, which academic economists missed almost entirely until the 1930's.[7] Keynesian parallels should be obvious, and also "structuralist" ones, involving *inter alia* Leontief's own input-output system.

1. Division of the private economy into "investment" and "consumption" sectors has become commonplace in the post-Keynesian generation, but it apparently dates from Marx.

2. If I am justified in including certain imprecisely specified "functions"—Section IV, equations (15)-(16) of the Appendix—into the Marxian schema,[8] he was an embryonic general-equilibrium theorist in advance of Leon Walras.

3. Marx presents a theory of underemployment equilibrium well in advance of Keynes, with the unemployment rate tending, for structural reasons, to increase over time.

4. The notion of a minimum rate of profit, below which capitalists will seek to hoard their savings in monetary form, seems to be a first cousin to the Keynesian liquidity trap in interest theory. Indeed, Marx's entire interest theory concentrates on equalizing returns to "money" and other capital; it may be a modern monetary one ahead of its time,[9] although my old-fashioned inclination is to believe otherwise on balance.

5. Marx antedated current institutionalist and structuralist writers, from Veblen and Ayres to Leontief and Chenery, in downgrading the importance of prices, and price-induced substitutions, as compared with purely technical production relations.[10]

It is the unimportance of prices, and *a fortiori* the unimportance of

[7] On the Marx-Keynes relationship, which became apparent at the end of this decade, compare Mark Blaug, *Economic Theory in Retrospect* (Irwin, 1962), p. 270 f.: "Most authors are impressed by the similarities . . . : Two-way disaggregation on the product side of the social accounts; a monetary theory of the rate of interest; the rejection of Say's Law; emphasis on the declining marginal efficiency of capital; and a chronic tendency toward oversaving in a mature economy."

[8] The Marxian underconsumptionists, such as Rosa Luxemburg, would doubtless deny the authenticity of (15). Other Marxists would also deny, with some horror, the entire notion of Marx as an "equilibrium" economist, since the term has taken on optimal and/ or full employment overtones above and beyond its service as a check on logical consistency. Compare Bronfenbrenner, "Classical and Marxian Macro-Economics," *op. cit.,* p. 150.

[9] For this interpretation, see Blaug, *op. cit.,* p. 265, citing *Das Kapital,* Vol. III, Chap. 13. But suppose, with, e.g., Böhm-Bawerk, an economy in which goods are borrowed and lent *in natura,* or in which the numeraire is an abstract, noncirculating unit of account. Would not a rate of interest prevail here too (under capitalism), along the lines of the exploitation theory of interest more commonly ascribed to Marx?

[10] Marx may himself have been anticipated by Ricardo in this respect, if one accepts Piero Sraffa's interpretation of *Production of Commodities by Means of Commodities* (Cambridge Univ. Press, 1960) as modernized Ricardianism. To me, the Sraffa system appears to omit the considerations underlying the Ricardian theory of rent. This makes it, as a Ricardian system, "Hamlet without the Dane," while, as a Marxian one, it is only "Hamlet without Rosencrantz and Guildenstern."

their divergence from values, for anything but the statical equilibration of markets and profit rates, that makes the labor theory of value so easy to uphold in the Marxian framework, both definitionally and as a workable approximation to competitive microeconomic facts. The point may be worth repeating: Nothing in Marx's aggregative "laws of motion of capitalism" would be affected in any significant way by any change in the pattern of divergences between prices and values, the p_i terms in the Appendix.

6. I owe to Leontief's 1937 paper to this Association an appreciation of the indebtedness to Marx of business cycle theory, which may itself be in something of an eclipse at the moment. A running quotation will touch the high spots of Leontief's appreciation:

Present-day business cycle analysis is clearly indebted to Marxian economics. It would hardly be an exaggeration to say that the three volumes of *Capital* helped more than any other single work to bring the whole problem into the forefront of economic discussion.

It is rather difficult to say how much Marx actually contributed to the solution of the problem. The two principal variants of the Marxian explanation of "economic crises" are well known. One is the theory of underinvestment, the other is the theory of underconsumption. Both might contain some grain of truth.

It is easy to find numerous hints and suggestions which can be interpreted as anticipating [each] and every modern theoretical construction.

[Here Leontief quotes from the Marx-Engels correspondence a passage indicating that "toward the end of his life Marx actually anticipated the statistical, mathematical approach to business cycle analysis."]

The significance of Marxian economics for modern business cycle theory lies, however, in the famous Marxian schemes of capital reproduction. An intelligent discussion of economic fluctuations must be based on some theoretical model revealing the fundamental structural characteristic of the existing economic system. In this field the original contributions of post-Marxian economics are rather uncertain. [Marx] developed the fundamental scheme describing the inter-relation between consumer and capital goods industries. The Marxian scheme still constitutes one of the few propositions concerning which there seems to exist a tolerable agreement among the majority of business cycle theorists.[11]

7. As in business fluctuations in contrast with stationary states (or Von Neumann rays!), so in imperfect in contrast with pure competition, Marx gives us no finished theory but an urgent sense of general unease, integrating facts and analysis, which has come to fruition long after his death. I remember my teacher, Frank H. Knight, warning me against undue interest in imperfect competition; specialists in that area, he said, usually ended up as Marxists!

8. Passing to more general and methodological matters, one hesitates to point out the smooth and natural articulation of Marxian statics and dynamics, because this virtue is shared with Marx's classical predecessors. However, *Das Kapital* was the last system with this feature—at least until Schumpeter. Static analysis took over the field in the 1870's, and we are not yet back to the Marxian level.

9. In the same way, Marx's assimilation of theory and practice, of economics and other social studies, is not new. He stands last, and pos-

sibly greatest, in a series from Locke through Hume and Smith, Ricardo and Mill, in what we self-consciously call today an interdisciplinary tradition. After Marx, such architectonics went out of fashion among economists, and was left to philosophers and sociologists uninterested in economics, or sated with it. It is characteristic that Keynes confined the "social philosophy" of the *General Theory* so largely into one chapter. Following his (and Schumpeter's) day, economists are once more raising their sights to embrace the other social disciplines, but no practitioner of Marx's own stature has yet emerged.

V

Because my critique of Marx differs both from the standard Böhm-Bawerkian textbook one and from the one Professor Samuelson is presenting today, let us consider it, if only as the reverse side of the appreciation expressed up to this point.

As for the statics, the besetting sin is ambiguity, a misdemeanor rather than a felony. Some of this ambiguity—for example, the frequent confusion between stocks and flows, particularly as regards constant capital—Marx might well have corrected had he lived to polish his system for a second edition. Another type of ambiguity, exemplified by the question of whether he proposed to set up a general equilibrium or disequilibrium system, cannot be resolved short of presenting the question to Marx's ghost, since it was not presented to him in the flesh. Rather, what I have in mind is the so-called "transformation problem," or the relation between values and prices.

Here the problem is less that Marx failed to make his meaning clear than that he offers alternative solutions with no basis for choice among them. If we accept my device (in the Appendix) of using p-coefficients as ratios of price to value (pure numbers) instead of absolute prices, one may argue, with textual justification, for some such equation as (3), which makes total and average values equal "total" and average prices. But one can argue just as readily, and again with textual justification, for making the sum of surplus values equal to the sum of profits (with all receipts and costs converted into prices). Some have also suggested setting price arbitrarily equal to value for one or another sector of the economy, which Marx never did. (A "luxury good" sector, composed of capitalists' consumption goods, is a common choice, because it does not reflect back to any other sector in the form of cost.) The point is not only that Marx made no clear choice, but that his system includes no clue for making one. The system works equally well either way, but with, in general, different results.[12] One is

[12] For a demonstration that the results do in fact differ, with a three-sector model, so that no single set of prices satisfies both of Marx's "invariance criteria," see Blaug, *op. cit.*, pp. 213-15 (correcting several misprints).

reminded of Mrs. Robinson's strictures against neoclassical price theory and its ambiguous treatment of "normal profits."[13] The "transformation problem" is the Marxian equivalent.

Allied to this ambiguity is another, involved in System II and equation (9) of the Appendix. This is the aggregative equality of supply and demand. Should it be expressed in terms of value (labor time) or of price (labor time as modified by p-coefficients)? Since supply and demand are market phenomena, and hence involve market prices, I have chosen the second alternative, following a suggestion by Mr. Yutaka Kosai. Most of the Marxian illustrations run in value terms, however, as did my own earlier efforts. Clearly, a substantive difference is involved, except in the trivial special case where all p-coefficients are equal to unity.

VI

Passing to the Marxian dynamics, I have somewhat less to add to the standard bourgeois appraisals. But once again, it is not completely clear what Marx is saying. Is the motive force of capitalist decline a falling rate of profit plus a liquidity trap of some sort, a tendency toward overproduction and underconsumption manifest in "realization crises," or some dilemma compounded of the two? There is again a related ambiguity: is collapse to come more or less automatically from accumulated disgust with prolonged stagnation and increasing unemployment, or can we expect the revolution before such a point is reached? My own interpretation leans toward a "dilemma model," with the realization crisis the dominant weakness, insofar as monopoly or oligopoly may prop up the profit rate by raising the rate of exploitation for a capitalist class which forms a diminishing proportion of the population.[14] On the issue of stagnation versus cataclysm, or the timing of the revolution, I am not sure Marx ever made up his mind, after disappointment of his hopes for the late 1840's. He would take his socialist revolution either way and at any time, and the sooner the better!

---

[13] Joan Robinson, "The Basic Theory of Normal Prices," *Q.J.E.*, Feb., 1962, pp. 10-12.

[14] I owe to Professor Nobuo Okishio an interpretation of the falling rate of profit .which would, if valid, apply under monopolistic as well as competitive conditions. Ignoring all distinctions between stocks and flows by setting our d equal to unity, we have:

$$P' = \frac{S}{C+V} < \frac{S+V}{C} = \frac{\text{"living labor"}}{\text{"dead labor"}}$$

In the Marxian vision of technical progress (*Das Kapital*, Vol. III, Chaps. 4–6, 13–15), by Okishio's interpretation, it is really this last ratio rather than the organic composition of capital k, which tends to fall over time. (As has been remarked frequently, especially by students of Chap. 14, Marx was less dogmatic about "Marx's Law" than many of his followers have been.) Let us agree that the living-labor/dead-labor ratio falls over time, but this ratio is clearly greater than the rate of profit itself. It does not follow that P' falls over time, since the fall of the capital-labor ratio could be counteracted by a rise in the ratio $SC/[(C+V)(S+V)]$.

Assuming these ambiguities resolved, the principal shortcoming of Marx's uniquely original and influential dynamics appears to be inter-disciplinary—an odd weakness indeed, in view of Marx's own stress on the unity of the social studies, history, and philosophy. Two illustrations involve logic and political theory, respectively.

As regards logic, I can do no better than repeat the main point of Professor Murray Wolfson's recent logical-positivist critique.[15] To Wolfson, Marx's forecast of capitalist downfall is so imprecise as to time, place, and pattern, that it is difficult to imagine any sequence of historical events in finite time as refuting it. Being irrefutable, the Marxian dynamics become, by logical-positivist criteria at any rate, essentially meaningless.[16] And indeed, it seems as difficult to cite Russian or Chinese semicapitalist or developing-country experience as supporting the Marxian system as to cite the last century of American or Western European advanced-capitalist experience as disconfirming it more than temporarily.

In the domain of political philosophy, Marx's theory of the state and its economic functions, however revolutionary in its own day, seems fundamentally outmoded in the large by institutional developments. It is no longer enough to laugh off the capitalist state as "merely" the instrument of the capitalist ruling class. Even accepting this proposition with fewer reservations than most Americans do, its significance is no longer so obvious as it was in Marx's lifetime. Viewed purely as an instrument of the capitalist class, the state has an interest in preserving the capitalist order, and need not sit idly by on bayonets while that order crumbles away in depression and stagnation. Furthermore, the modern state commands resources of monetary and fiscal policy undreamed of in Marx's philosophy, which was apparently shackled to metallism and budgetary balance by the implications of the labor theory of value. Whatever the flaws of contemporary mone-tary-fiscal economics, it will not do to dismiss them as "creation of fictitious values," an orthodox Marxist procedure of the New Deal period.[17]

The Soviet trend toward "competitive coexistence" since Stalin's

[15] Murray Wolfson, *A Reappraisal of Marxian Economics* (Columbia Univ. Press, 1966).

[16] It is probable that the younger Marx and Engels, in the halcyon days of the *Communist Manifesto* (1848) did indeed anticipate a more or less immediate collapse of capitalism, beginning in the advanced areas of Western Europe. In this interpretation, Marx stands disconfirmed, but only in a preliminary or juvenile version which anticipates *Das Kapital* by twenty years and more.

[17] Neo-Marxists (revisionists?) of that period were more perceptive, as witness Rogin, "Marxian Economics and Government Policy," *op. cit.*, p. 14: "Marx never envisaged state action on a large scale in the interest of the masses, of recovery, and of economic stabilization. In fact, the main task of those who wish to employ the Marxian theory in concrete economic analysis is to adapt it to the requirements of an economic process which involves a vast amount of government regulation and participation."

death is often associated with retreat from Marx's original position, and the Chinese charge of "modern revisionism" is entirely plausible. Instead of denying the efficacy of monetary and fiscal policy in averting stagnation, the revisionist line calls the method wasteful, bellicose, and divorced from the people's welfare as compared with all-out "rational" socialist planning for growth and progress. Whatever one may think of this argument—to me, the issue remains wide open—it has progressed a long way from any volume of *Das Kapital*.

VII

The foregoing estimate, viewing Marxism primarily as macroeconomics, is intended as neither outright acceptance nor outright rejection. On the static side, it is probably less unsympathetic than most American classroom presentations. On the dynamic side, it is more conventionally critical, but not to the point of suggesting that Marxian dynamics is completely outmoded, no longer worth taking seriously, or an impossible basis for useful extensions.

Let me close by repeating another position I have already taken.[18] I look forward from the centenary of *Das Kapital* to the time when, in America as elsewhere, serious academic work in controversial Marxian economics is carried on by professed Marxian economists as well as others like myself, and when the ideological handicap under which Marxists presently suffer in seeking academic preferment is lowered from three strikes to one, and preferably abandoned completely. And, if the point needs making before this audience, I also look forward (with considerably less confidence) to similar freedom for controversial bourgeois economics by bourgeois economists in predominantly socialist countries.

[18] M. Bronfenbrenner, "Marxian Economics in the United States," *A.E.R.*, Dec., 1964.

APPENDIX

Notation

Department I (subscript 1), produces capital goods.

Department II (subscript 2), produces consumption goods.

W—Value, measured in labor-hours (hours of socially-necessary labor).

C—Constant capital, depreciation and intermediate goods, measured in labor-hours (a flow, not a stock).

V—Variable capital, wages of production workers, measured in labor-hours (of product, not of actual labor).*

S—Surplus value, property income plus salaries, measured in labor-hours.

p—Ratio of price to value, a pure number.

*V_o—The "full employment" value of $\sum V_i$.

w—Wage rate of productive labor, measured in labor-hours of product.

S'—Rate of surplus value, S/V.

P'—Rate of profit, essentially $S/(C+V)$.

K—Fixed capital (a stock).

d—Depreciation rate, essentially C/K.

g—Proportion of S invested in output of Department I.

k—Organic composition of capital, C/V.

h—Capital coefficient of Department II, W_1/W_2.

u—Rate of unemployment, $1-(\sum V_i/V_o)$.

Equation Systems

I. Labor Theory of Value (8 Equations)

1–2. $W_i = C_i + V_i + S_i \quad (i = 1, 2)$

3. $\sum W_i = \sum p_i W_i$

4–5. $w = \dfrac{V_i}{S_i + V_i}$

6. $S' = \dfrac{1 - w}{w}$

7–8. $P' = \dfrac{S_i p_i}{V_i + (C_i/d_i)} \quad \text{or} \quad P' = \dfrac{S_i' p_i}{1 + (k_i/d_i)}$

Notes:

1. From (4)–(6) we also have $1/w = S' + 1$.

2. The wage rate w is also constrained by the standard of living, expressed by the past wage rate w_{-1}. This constraint does not take equational form; it may be expressed by the condition that the quotient or difference of w and w_{-1} should not exceed some constant ϵ in difference from unity or in absolute value, respectively.

II. Supply and Demand (1 Equation)

9. $p_1[V_1 + (1 - g)S_1] = p_2(C_2 + gS_2)$

III. Structural Equations and Identities (4 Equations)

10–11. $k_i = \dfrac{C_i}{V_i}$

12. $\sum C_i = \sum d_i K_i$

13. $u = 1 - \dfrac{\sum V_i}{V_o}$

14. $h = \dfrac{W_1}{W_2}$

IV. Functional Relationships (2 Equations)

   15. $g = g(P', h, W_2) \cdots$ all derivatives nonnegative.

   16. $u = u(S'_{-1})$, so that, from (7–8) $P' = \dfrac{[u(S'_{-1})]^{-1}}{1 + (k/d)}$
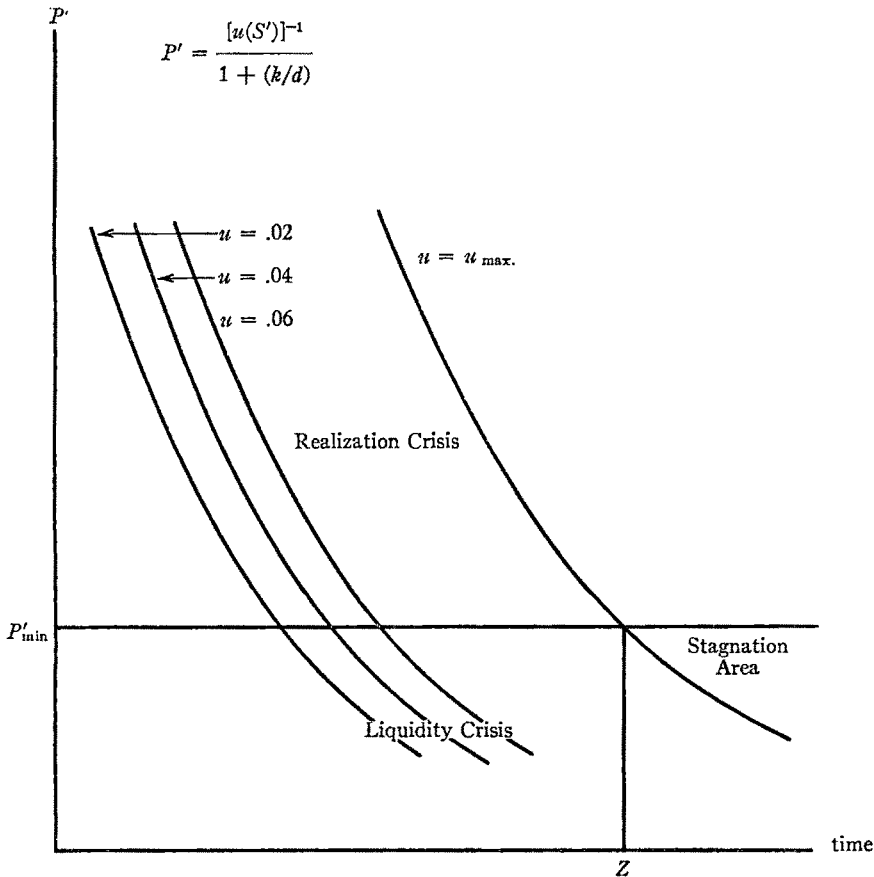
   Note: S'_{-1} in (16) is a proxy for the recent-past rate of surplus value, and does not refer exclusively to the period immediately preceding. The derivative du/dS_{-1} should be considered positive.

Unknowns (16)

   W_i, C_i, V_i, S_i, p_i—10 in all
   S', P', w, h, g, u—6 in all

   Note: All other variables are technologically determined.

"Dilemma" Diagram

$$P' = \frac{[u(S')]^{-1}}{1 + (k/d)}$$

$u = .02$

$u = .04$

$u = .06$

$u = u_{\text{max.}}$

Realization Crisis

P'_{min}

Stagnation Area

Liquidity Crisis

Z

time

Notes: 1. k presumed to increase with time.
          2. No time trends in d or P'_{min}.
          3. $Z = Zusammenbruch$ (collapse, breakdown).

DISCUSSION

EVSEY D. DOMAR: My comments are directed at Bronfenbrenner and Erlich's papers, which arrived last week, rather than at Samuelson's, which made its appearance yesterday.

In the spirit of the "birthday party," I am willing to accept all contributions of Marx listed by Professor Bronfenbrenner, and to share Professor Erlich's enthusiasm for the two-sector model derived from Marx's scheme of expanded reproduction. Such a model is relevant to economic planning in countries where the limited capacity of the capital goods industry is a major obstacle to growth, as was indeed the case in the Soviet Union in the 1920's and 1930's, and as is true of many underdeveloped countries today. Though G. A. Feldman is known as the father of such models, his contemporary, E. Preobrazhensky, suggested, though only implicitly, a more interesting one in 1926. With the recent translation of his book, *The New Economics*, into English (Oxford, Clarendon Press, 1965), the popularity of such models is likely to rise. For that matter, they are already quite popular, particularly among M.I.T. graduate students. I cannot assert that all two-sector models stem from Marx, but many must have been inspired by him, even if indirectly.

But as I look over the list of Marx's contributions presented by the two speakers, I cannot escape the feeling that something important was left out. It is as though I had a description of some missing person in terms of height, weight, color of eyes and hair, etc., yet was unable to picture the person.

As you know, Leo Tolstoi sought true wisdom among Russian peasants. There are no peasants in this country; hence we appeal to the "man in the street" or to the "intelligent layman." If asked what he associated with Marx, the layman would mention revolution, communism, and perhaps even forced industrialization. And he would be right. For Marx was no ordinary economist doing research for fun, glory, or money. He was a revolutionary whose main purpose in life was the building of a "scientific" foundation for the overthrow of capitalism. After all, most of his important ideas had already appeared in the *Communist Manifesto*. It was for a good reason that Schumpeter titled his first chapter on Marx, "Marx the Prophet."

In Marx's scheme, four causes could have spelled the doom of capitalism. First, there was the general economic evolution. When an economic system cannot make full use of its productive potential, it is replaced by a superior system that can. Marx's theory of economic evolution is, in my opinion, one of his greatest contributions. (Bronfenbrenner, who calls Marx "the greatest social scientist of all time," will probably agree with me; I only wish he had pursued the idea further.) But as a revolutionary doctrine it simply will not do.

Next, capitalism could break down because of the shortage of demand created by the lopsided distribution of income. This thought appears in Marx quite frequently, but most authorities, including Engels, deny that Marx was an underconsumptionist. Perhaps Marx's intuition warned him that this problem could easily be solved.

The last two causes are the declining rate of profit and the increasing misery of the workers. The first would produce capitalist crises of ever increasing magnitudes; the second, partly following from the first, would eventually lead the workers to revolt and overthrow capitalism.

But Erlich dismisses the falling rate of profit as "a very weak rod to lean upon," and approvingly refers to Mrs. Robinson, who had said that "[Marx's] explanation of the falling tendency of profits explains nothing at all." (*An Essay on Marxian Economics*, p. 42.) All that remains now is the increasing misery of the workers. And now Erlich tells us that Marx did not mean that at all.

Whether Erlich is right or wrong depends, as in other index-number problems, on the weights to be attached to Marx's differing (and even contradictory) statements. With Erlich's weights, his result presumably follows. And so Marx's revolutionary doctrine is completely demolished. There is nothing left. The familiar prayer seeking divine protection against one's friends is now in order.

That Marxism is first and foremost a revolutionary doctrine can be confirmed by examining its social impact: very modest in advanced countries (and with a negative trend at that) and tremendous in the underdeveloped ones. Of course Marxism played an enormous role in the Russian revolution, but I have often wondered whether it helped or handicapped the communists after they took over. Granted their objectives, I think it helped them: the policy of rapid industrialization by means of great accumulation of capital and by the transfer of masses of peasants into industry is easily deduced from Marx. But now that the stage of "primary capital accumulation" is over, further progress of the Soviet economy depends on efficient allocation of resources; here there is practically nothing that the Russians can take from Marx. And so they will continue to venerate him in word and disregard him in deed, as they manage their economy by means of prices, interest, and profits— a process that has already begun. That he suffered the common fate of prophets, Marx should be the last to complain: he had always asserted the supremacy of economic forces over ideas.

DONALD F. GORDON: Needless to say, I cannot, in a few minutes, comment upon more than a small fraction of the points raised in these three appraisals of Marxian economics. Rather than treat the papers individually, it will be more convenient to raise two or three main issues, and the bearing of my comments upon these papers will be in part implicit.

First the labor theory of value. I could not agree more with Professor Samuelson's Proposition 2, and further that (as he has elsewhere phrased it) the definition of value by labor seems to have been an unnecessary detour in an attempt to understand modern capitalism, whatever one's appraisal of the latter. While this is true *ex post*, I think we would be more sympathetic to the Ricardo-Marx definitions if we look at it *ex ante*. Professor Samuelson has somewhere said that while we cannot prevent national income statisticians from adding up totals of private and public balance sheets, and we cannot pass a law against calling the total "capital," we should realize that this number is

not something we can kick or lean against. Similarly, we add up certain flows, and after some index number adjustments, we call the total real net national product. I am inclined to think that we kick and lean against this total a little too often. How much more satisfactory it would be in principle if we could bring ourselves to believe in a national aggregate of interpersonal and cardinal utils—something which Jevons appears to have believed would be produced by progress in the science of biology.

However that may be, when Ricardo wished to discuss the distribution of an aggregate flow of something, he had neither net national utils nor net national product, since he lacked (and in fact specifically rejected) the notion of an index number. Perhaps he should have stayed with the one-commodity world of the *Essay on Profits*. In fact, he chose to add up the labor-hours required to produce each unit of each good and to discuss the distribution of this total, fully understanding that relative exchange rates between goods are not equal to the ratios of these absolute numbers. This is admittedly highly unsatisfactory; yet we must remember that the rationale in the foundations of national income accounting cannot itself stand too close an inspection.

I believe we must interpret Marx as accepting Ricardo's definition despite the attempt that he made in the opening pages of Volume I to deduce the definition as a theorem by what is a play on the word, or rather the symbol, for equality. It follows that by definition all laborers, including Miss Elizabeth Taylor, are exploited unless they receive the whole of the national product. Despite this perverse outcome, the intellectual (not ideological) roots of the labor theory of absolute value lay in a very real problem. I doubt if, apart from this problem, any grown man would have carried a labor theory of any kind past the two pages of Smith's early and rude state of society. If Ricardo considered the distribution of something to be the major question of political economy, he had to have a fairly definite, even if unsatisfactory, idea of what he was distributing.

My next comment concerns the celebrated accumulation-breakdown controversy. In this I am inclined to side with Professor Erlich's search for the key to Marxian historical dynamics in Volume I rather than Professor Bronfenbrenner's preference for Volume III. The Volume I interpretation of the capitalist breakdown has been obscured by a large literature treating the collapse in terms of unemployment, declining rates of profit, realization and disproportionality crises, and in the case of Bronfenbrenner, liquidity preference. I would not wish to deny that with some effort one may be able to construct a breakdown out of Volume III and certain parts of the *Theories of Surplus Value*. What I would argue (with Mrs. Robinson) is that a breakdown comes forth less artificially and with less strain from Volume I, and it is a breakdown that is more consistent with Marx's fundamental Ricardian bias toward the question of aggregate demand.

The first point to notice is that if the breakdown is described anywhere in *Kapital* it is in Volume I and specifically in Chapter 32. It is natural to suppose (and I get the distinct impression) that in this dramatic chapter Marx intends, by drawing together the strands of argument that he has been developing earlier, to point to an inevitable conclusion. Putting together other

materials with Chapter 24, upon which Professor Erlich has properly laid emphasis, a fairly simple explanation of the collapse occurs.

Induced and autonomous laborsaving innovations, together with a highly elastic supply curve of capital, lower wages—either absolutely or as a share of the national income. Economies of scale drive out small firms, and small capitalists are driven into the proletariat. (The latter proposition presumably requires a capital market sufficiently imperfect to prevent small capitalists from uniting to form General Motors, but this implicit assumption is common in many orthodox pre-Marxians.) Thus the national income becomes increasingly polarized, the limit of this process arriving when one individual, or at least one family, receives all the property income, and this constituting a constantly increasing share of all income. Naturally this is politically unstable—an instability enhanced by capitalism's perverse practice of geographically drawing the proletariat together and thus enhancing class consciousness. Jay Gould to the contrary notwithstanding, one cannot hire half the working class to suppress the other half, and the revolution occurs. Whether it is violent or peaceful becomes an essentially trivial question, which may explain Marx's ambiguity on this point.

Perhaps this explanation is too simple since it ignores the reserve army, the declining rate of profit, and any mention of underconsumption. One explanation (Schumpeter's) of the reserve army is in terms of the continuous disequilibrium created by constant technical change. An alternative explanation could be that, following Samuelson, production functions consist of finite numbers of alternative but fixed coefficients, which would create at least limited opportunities for unemployment if substitution possibilities through consumption were sufficiently restricted.

I think it is important to note that this revolution occurs in *Das Kapital* hundreds of pages before there is any mention of the specific Marxian explanation for the declining rate of profit. Moreover, this famous law which plays such a large part in post-Marxian discussion occupies a very brief part of Marx's thousands of pages of economic analysis. The few lines which support an underconsumptionist position are specifically rejected elsewhere. Who knows but that Marx might have modified or conceivably thrown away either or both of these ideas had he and not Engels published Volume III. In any case, the Volume I analysis can stand by itself.

Of course, as an empirical matter Marx was wrong on almost all points. (Needless to say, this does not bear on the desirability of nationalization or planning.) This leads me to share Professor Samuelson's view that, as an economist, he was a minor post-Ricardian, and we should remember that overwhelmingly the largest part of his intellectual work was in economics, not other social sciences. The interesting question therefore may be: Why does half the world regard him as a patron saint? Why in particular do we witness the ironic spectacle of the underdeveloped world embracing Marx and socialism so unreservedly, while we know that in his praise for the productive powers of capitalism Marx had few if any peers? Far from being hidden in an obscure passage of a weighty tome, this praise was trumpeted in the *Communist Manifesto* which people do, in fact, read. Such questions must be answered, if

they can be, by other disciplines. Are there perhaps at this moment other meetings of these other disciplines appraising Marx and discussing these questions? Perhaps we should be there and not here.

H. SCOTT GORDON: Only one of the three papers which have been presented, Professor Bronfenbrenner's, was available early enough for me to study it in preparation for this meeting, so I will confine my remarks to some comments on this paper and follow with some general comments of my own on *Das Kapital.*

The last half of Bronfenbrenner's paper, as he himself notes, is to be viewed in conjunction with his interpretation of Marxian macroeconomics in terms of a general equilibrium model which he has earlier advanced in published papers. This is a neat and clear static model which, in its latest version, consists of fifteen unknowns and fifteen equations. Part of the model is rendered dynamic in order to generate the "contradictions of capitalism" which is the essential element of Marx's view of capitalism as an inherently transitory system of economic and social relationships.

Concerning the model itself, I find it difficult to say much more than that it is a possible interpretation of *Das Kapital.* All but two of the equations of which it is composed are definitional statements or constants. These thirteen are authentically Marxian, consisting of constructs which Marx makes extensive explicit use of in his analysis. The remaining two equations are general functional relationships which, as Bronfenbrenner himself admits, are not to be found explicitly in *Das Kapital,* but have to be distilled from Marx's verbal and numerical analysis, much of which is unclear and ambiguous. If one holds the methodological view that the vital part of an economic model consists in the functional relationships that are built into it, all the rest being so much structural carpentry, then Bronfenbrenner's model, or rather Marx's system construed according to Bronfenbrenner's model, appears to be rather anemic. In the dynamics, only one functional statement enters, that there is a positive relationship between the rate of surplus value and the degree of unemployment. Now this statement, which necessarily implies that the rate of profit rises in recession and falls in prosperity, would be even more questionable empirically than the celebrated Marxian prediction about the secular course of the profit rate. But, in fact, the relation between the rate of surplus value and the degree of unemployment is left so ambiguous in *Das Kapital* that one can run the causal nexus of this relationship in either direction or, as Bronfenbrenner does, in both at once.

It is not clear from Bronfenbrenner's paper the extent to which his catalogue of "modern" elements in *Das Kapital* is a derivative of his interpretive model, though I think he must have gone beyond this model in deriving some of the elements in his list. Considering this catalogue as such, however, my own reaction to it, placed against a fresh reading of *Das Kapital,* is that Bronfenbrenner is, to say the least, rather generous to Marx. The "embryonic general equilibrium theory" is more in Bronfenbrenner's mind than it was in Marx's. The ideas of unemployment equilibrium and stagnation, liquidity trap, and business cycle theory are such ghostly suggestions in *Das Kapital*

that they are not really visible to a student of the history of economic thought who approaches the book with the agnosticism that is a necessary attribute of that discipline. I would not have thought that something which is no more than "an urgent sense of general unease" about imperfect competition can be recorded as a seminal anticipation of modern work in this field. I do not find the "smooth and natural articulation of . . . statics and dynamics" of which Bronfenbrenner speaks, unless he is referring not to Marxian economics but to general Marxian historical theory, which hardly appears at all in *Das Kapital*. In short, I would quarrel seriously with all of the elements of Bronfenbrenner's catalogue except the first, which refers to Marx's development in the latter part of Volume II, of a two-sector analysis of economic processes, one sector being the investment goods industries and the other the consumption goods industries, and I would have put another item, which Bronfenbrenner does not mention—Marx's conception of the economy as a set of interrelated circular flow processes—at the head of the list.

Despite the dangers of overimaginative interpretation, one cannot in trying to come to grips with Marx, follow the excellent injunction of Engels in his preface to Volume III, that one "should above all learn to read the works he wishes to use just as the author had written them, and above all without reading into them that [which] they do not contain." *Das Kapital* is not such a simple and straightforward work as to permit such a literal reading. But it would be a mistake, in my opinion, to approach it solely in terms of modern analytical economics, for a fundamental purpose of the book is to do something that modern economists rule out of economics altogether—the making of normative judgments. *Das Kapital*, it is true, has to do with the structural-functional mechanism of commodity production under capitalism, which is a matter for "positive" economics, but even more so it is a book about justice. There is surprisingly little in it, in fact, on the contradictions of capitalism or the mechanism of breakdown, but the view of capitalism as a system of exploitation pervades the book generally and this is especially true of the only volume of it that Marx himself completed.

The deep mystery that surrounds *Das Kapital*, however, is not really its content, but its influence. How did such a long, verbose, abstract, tedious, badly-written, difficult labyrinth of a book become the Talmud and Koran for half the world? Now that is a very big and possibly an unanswerable question, but centenaries are legitimate occasions for asking big questions. Marxism is a "great fact" of the twentieth century. That it should have become so is perhaps more of a mystery to a person trained in Western economics than to others. But it is undeniably a faith that is moving mountains. In attempting to understand its leverage, we can probably get more assistance from Durkheim and Frazer and Freud, than from Walras and Keynes, for we are faced here with the influence that such things as totem and myth and fantasy have on human thought and action, not with anything that is as severely rational as modern economics.

THE EFFICIENCY OF EDUCATION IN ECONOMICS

EXPERIMENTS IN THE TEACHING OF BASIC ECONOMICS

By BERNARD F. HALEY
University of California, Santa Cruz

In March, 1965, the American Economic Association's Committee on Economic Education invited the submission of information about experimental courses in elementary economics together with, if possible, evaluations of the degree of success attained in achieving the objectives of the courses. The primary purpose was to make information about any such experiments available to other institutions and to the profession at large; the present paper is a shorter version of an analysis of the responses to the Committee's invitation. (The longer report is available from the Joint Council on Economic Education.)

About thirty institutions responded; but many of the submissions concerned experiments with particular techniques rather than whole experimental courses. Both kinds of information, however, should be of interest to the profession; consequently the present survey covers a somewhat broader terrain than the one defined by the Committee when it invited submissions.

In the first part of the present survey experiments with the content of the introductory course are reviewed; in the second part, experiments with different techniques in teaching the course are examined.

I. *Different Types of Introductory Course*

There is considerable experimentation currently with types of introductory course that differ in greater or less degree from the traditional course. To consider first the more conservative variations, perhaps the least radical has been the offering of two alternative principles courses, or alternative sections of the same course, in one of which the approach is essentially nonmathematical, while in the other the theoretical analysis is presented mathematically. Examples are to be found at Marquette, Stanford, Williams, Wesleyan, and Yale.

A quite different sort of development has been the experimentation with a one-term principles course, sufficiently self-contained to be taken as a terminal course, designed for the student who does not have room in his program for a two-term course. Both Carnegie Institute of Technology and Wesleyan University offer such a course, complement-

ed by a second one-term course for those students who, it may turn out, in fact want to complete the sequence. In the first term of the Carnegie Institute course, allocation theory is developed at a very general level while income theory and analysis are more thoroughly treated. A restricted sample of problem areas is explored. In the Wesleyan experimental course the stress is on principles with a view to encouraging students to go on to more specialized problem-area courses after one term. The Carnegie Institute experiment is of particular interest because of the thorough evaluation and testing to which the course was subjected.[1]

Still another modification of the standard principles course has been its adaptation to meet the needs of upper division students not majoring in economics who failed to take the sophomore course. An example is to be found at the University of Minnesota where both a sophomore and an upper division course are offered.

Probably the most common sort of experimentation with course content has been with an essentially institutional and/or problem-oriented approach, aimed primarily at the student not expected to take further work in economics and consequently avoiding to the fullest extent possible the exegesis of refined and technical theoretical analysis.

One example is to be found at Hamilton College. Broad economic principles are introduced to the student in connection with the examination of a number of problem areas; but graphs and equations, statistics, T-accounts, multiplier and elasticity formulas, and other technical matters are avoided. A somewhat similar one-term course has recently been developed at the University of Pittsburgh; and at Harpur College there was experimentation for several years with a problem-oriented, institutional course in which a considerable amount of economic history and history of economic thought was interwoven.[2]

In some cases experimentation has been devoted to a second and different type of introductory course while the traditional principles course continues to be offered. Duke University, for example, is experimenting with a new full-year freshman course while retaining the traditional sophomore principles course for the present at least. In the new course the first term is devoted mainly to allocation and income theory while the second term, with some slight similarity to the Harpur College course, includes a considerable amount of economic history and the history of economic thought. Vassar College also offers an introductory course, mainly for freshmen, institutional rather than

[1] The syllabus of the course (C841) and/or the *Evaluation* of the course may be requested from Professor Phillip Saunders, Department of Economics, Carnegie Institute of Technology, Pittsburgh 13, Pennsylvania.

[2] Professor Laurence Leamer, Chairman, Department of Economics, Harpur College, Binghamton, New York, 13901, will provide a full description of the course on request.

theoretical, in addition to the regular principles course in the second year. Another example is a University of Minnesota one-term course "about economics" offered for students who would like a brief introduction to the subject. After an introductory treatment of the history of economic thought, the course surveys the principal branches of economics.

A course quite similar to the second semester of the Duke University freshman course has been offered as a full-year introductory course at the University of San Francisco since 1961. The principles of economics are taught through the medium of the history of economic thought, with economic history woven into the presentation. The standard principles course is also offered, primarily for nonmajor students.

A quite unique one-term course is offered at the University of Southern California specifically designed for upper division honors juniors and seniors in the sciences and humanities. It is not a principles course but is more like the Minnesota course "about economics."[3] With somewhat similar objectives, there is an upper division course at Harvard University, offered in addition to the usual principles course and taken particularly by science and engineering students. Under the title, "The Growth and Structure of the American Economy," the course has a conceptual framework provided by input-output analysis, and the approach throughout emphasizes empirical materials.[4]

A few conclusions may be drawn with regard to this wide variety of experiments with course content: (1) There is considerable experimentation with the level of theoretical analysis at which the introductory course is pitched. At one extreme are courses that offer a quite advanced theoretical treatment; at the other are courses primarily concerned with institutional structure, mainly problem-oriented, utilizing only very broad theoretical principles. (2) A number of interesting experiments are under way with courses integrating principles, history of economic thought, and economic history. (3) Quite a few institutions are experimenting with different types of introductory courses for meeting the needs of different types of student: the mathematical course, the one-term course for students with little room in their programs for economics, the introductory course for upper division students, the course for humanities and sciences honors students, the course primarily for engineering and science majors.

The only case, of all those reviewed, in which the experimentation has been adequately designed in advance, subjected to controls, and

[3] A description of the course may be obtained from Professor Spencer D. Pollard, Department of Economics, University of Southern California, University Park, Los Angeles, California, 90007.
[4] The reading list for the course will be sent on request. Address: Professor Wassily Leontief, Department of Economics, Harvard University, Cambridge, Massachusetts, 02138.

carried out in a way making possible the testing of a hypothesis has been the new sequence of courses at the Carnegie Institute of Technology.

II. *Teaching Techniques*

Reliance on large lecture classes has seemed to be a relatively harmless way of economizing on both budget and manpower allocated to the elementary course in many large institutions in a period in which graduate programs have become increasingly expensive, inflation has required substantial salary increases, and the pressure for research productivity has necessitated reductions in teaching loads. The tendency for class size to increase has to a degree been rationalized by acceptance of the results of numerous experiments designed to test the relative effectiveness of large and small classes. In general these have seemed to show that the principles course can be "taught" as effectively in classes of 150 as in classes of 20. It was easy then for administrators to argue that a lecture class of 300 is as likely to produce the desired results as one of 150; and a class of 600 as one of 300. So why repeat courses? A preferable alternative was to increase salaries or reduce teaching loads. Faculties agreed.

Now, however, we are not so sure. Elementary economics is a course, many of us believe, in which frequent and easy interchange between teacher and student is very important. As class sizes have increased this has become less and less possible until in many cases it has disappeared altogether. The student has become more and more a passive, note-taking auditor.

An awareness of the passivity generated by the lecture method is to a considerable extent responsible for a variety of experiments with techniques designed to induce the student to take a more active role in the learning process. Other experiments have been directed to the improvement of the lecture method itself.

Some of the more recently developed devices for increasing student participation in the learning process have been the tutorial, the workshop, the panel discussion, and the assignment of problems. Currently, at Wesleyan University, considerable use is made of "tutorials" by the different instructors: two or three times during the semester regular class meetings are discontinued for a week in favor of small group meetings to discuss student papers on an assigned topic. Vassar College has been making good use of workshops, in which students do assigned problems under the direction of senior major students; and "buzz sessions," which are voluntary meetings of small groups of students, organized and run by the students themselves, to discuss selected aspects of the course. Both techniques have been found effective

in increasing student interest and participation in the course. Eastern Michigan University has found that panel discussions by the students themselves have proved useful.

All of these means of stimulating student participation in the learning process have, in one form or another, been subjects of experimentation for some time. More recent developments are the experiments with programmed instruction, games and simulation, and television.

Programmed Instruction. For those concerned about the ineffectiveness of the large lecture class in the teaching of the elementary course, the emergence of programmed instruction as a teaching technique is particularly encouraging. Although the results of experiments to date are by no means decisive, they do give reason to hope that we may here have a means of developing the student's own initiative in the learning process, getting away from the lecture method for at least a substantial part of the course, and achieving both of these highly desirable objectives without increasing the burden on either staff or budget.

Rendigs Fels has been a pioneer in the application of programmed instruction to the principles course. During 1961-62 at Vanderbilt University he, together with D. R. Starleaf, conducted two controlled experiments designed to test the usefulness of programmed instruction in teaching theoretical concepts [1]. Their results were inconclusive. Similar experiments with Fels's version of programmed materials by D. R. Fusfeld and Gregory Jump [2] at the University of Michigan led to the conclusion that students who were given programmed material and excused from section meetings did just about as well as those who were required to attend section meetings but were not given programmed material. (All students attended the lectures and read the same text.)

At the University of New Hampshire in the spring semester of 1966 one section of the introductory course was taught supply and demand analysis with a standard text and lectures; another section relied exclusively on a programmed instruction text. The results were statistically inconclusive.

Further controlled experimentation with the technique has been carried out by Keith Lumsden at Stanford University and at the University of California, San Diego; and his results are reported in a companion paper at this session.

Now that there are programmed instruction texts available (and others in process of preparation), it is to be expected that this method will have wide use in 1966-67. Institutions (in addition to those already referred to) that are known to be planning to use it are: University of California, Santa Cruz; Vassar College; Wesleyan University;

University of Wisconsin. It is to be hoped that additional controlled experiments with its use by many different experimenters will occur, and that the results can be made known to the profession.

Games and Simulation. The use in instruction of a simulated market in which the students play the participating roles has been tried out since 1959 by M. L. Joseph, Carnegie Institute of Technology [3]. He has described role-playing experiments in market price determination, oligopoly pricing, and labor-management negotiations which he has used in the introductory course and the course in labor economics. In his judgment the experiments have been very helpful as teaching devices and are well worth the extra time they require.

A somewhat more elaborate game simulating the behavior of a simplified competitive economy was recently tried out by Herbert Fraser at Washington University with first- and second-year students.[5] Although no attempt was made to test the effectiveness of the technique, he and others associated with the experiment found the students to be more interested in the course than usual, and later student interviews showed that retention of what the students had learned was high.

The class is first introduced to a one-commodity two-factor Ricardian economy. Each student is given a piece of land to which is attached a revealed land-labor production function. He is instructed to hire workers to maximize the rent of his land or to hire out his own labor at the highest wage available, whichever would give him the higher income. Although at the beginning there is general ignorance of the equilibrium conditions, after a few sets of bids and offers are recorded, an approximation of the theoretical equilibrium situation is reached.

Even at this early stage of the game, the students learn for themselves the major principles underlying the allocation of resources in a competitive economy, how free markets tend to allocate resources optimally, and what are some major sources of income inequality.

The game proceeds through successive stages to take into account: (1) increase in population; (2) a government sector, with taxes and transfer payments; (3) a money of account; (4) the division of household income between consumption and saving; (5) a schedule of the marginal efficiency of investment on each farm, and lending and borrowing at a market rate of interest; (6) the rate of growth implicit in the rate of investment; (7) social versus private investment in relation to the rate of growth.

One other game for instructional purposes has been reported by John A. Carlson, Purdue University. Although his game has been designed for use in a course in macroeconomic analysis at the intermedi-

[5] Fraser, currently in Colombia, will provide a full report of the experiment on request when he becomes a member of the Department of Economics, Earlham College, next fall.

ate level, it might with some difficulty be adapted to the introductory course. The purpose is to provide the students of the class with an opportunity to discover for themselves what determines the level of net national product from period to period in a simplified, simulated, national economy.[6]

The principal advantage of games and simulation as a teaching device is that the students themselves participate in the discovery of principles and thus are likely to take a more active interest in the learning process than would otherwise be the case. The disadvantages are the lack of appropriate parallel text materials and the amount of time consumed in running through the game or simulation. However, Karl B. Radov, University of California, Irvine, reports that he is trying out currently a number of computer programs suggested by Fraser's game that should be less costly of time for both student and instructor.

Television. An important stimulus to the use of television in instruction was undoubtedly provided by the success of the nationally broadcast course, "The American Economy," cosponsored by the National Task Force on Economic Education, the Learning Resources Institute, the American Economic Association, and the Joint Council on Economic Education, and shown over 241 television stations in 1962-63.

Many local and regional uses have been made of open-circuit television for the teaching of the introductory economics course since 1963. One example is a course offered by the University of Virginia designed for teachers in the elementary and secondary schools. About 150 teachers in five communities took the course during 1965-66; and it will be repeated for a larger number of areas in 1966-67. Three-hour discussion meetings between the instructor and the various groups are held monthly. No attempt has been made to measure the effectiveness of the television course as compared with that of the in-service courses; but R. S. Johnson, who directs the television course, is satisfied that the students taking the televised course do about as well as those taking the regular course. The important advantage of the open-circuit course is that the University is thus able to reach more teachers in a single year than had been reached through in-service courses during the preceding two years.

For the college or university principles course, the greatest potential use of television, however, is at the institutional level—the use of closed-circuit, usually advance-videotaped television lectures in place of the lecturer-in-person for the large class. When thus used, the television medium can have all of the advantages of, and fewer disadvan-

[6] Further descriptive material will be provided on request. Address Professor John A. Carlson, Graduate School of Industrial Administration, Purdue University, Lafayette, Indiana, 47907.

tages than, the open-circuit course. The same careful preparation, use of visual aids, and introduction of special lecturers on opposite sides of controversial issues can be combined with better adaptation of the course of lectures to the level of instruction appropriate at the local level. The lectures can be viewed and heard simultaneously by students meeting in a number of smaller classrooms instead of in one great lecture hall. An audio intercommunication system between studio and classrooms (when the broadcast is "live") can be used to permit students' questions. More advantageously, discussion sections following the lectures can be utilized for obtaining student feedback, to introduce particularly timely materials or topics, and to give the student an opportunity to try out his views on policy issues and his ability in analysis. If budget permits, the lecturer can be provided with expert guidance in the use of the medium and the support of learning specialists, research assistants, and graphic artists.

At the University of Nebraska, C. R. McConnell reports, major reasons for the adoption of the new technique have been rapidly increasing enrollment, lack of the necessary number of full-time staff and graduate assistants to continue the course as a live lecture course, coupled with the necessity of scheduling principles classes at different hours. In 1965-66 about 820 students have heard and viewed McConnell's videotaped lectures in sections of about 50 each. McConnell is planning to test the effectiveness of the television course against that of the lecturer-in-person course in 1966-67.

Another institution which has experimented with televised instruction, primarily with a view to economizing on faculty required for teaching the elementary course, is the State University of New York at Buffalo. They continued the experiment over a five-year period, and finally abandoned the method because they concluded that the quality of instruction was less than could be achieved in large lecture classes.[7]

At the University of Tennessee, H. E. Jensen reports, the use of closed-circuit television for the lectures in the introductory course began in the fall of 1964. This is now a three-quarter course, and television is used for three lectures a week throughout the course. In the fall quarter 1966, 975 students attended the lectures in five classrooms. Students are encouraged to attend a two-hour tutorial session every week. In identical examinations given TV classes and non-TV classes, the TV classes have on the average performed slightly better than the conventional classes.

---

[7] Pennsylvania State University experimented with televised instruction in elementary economics in 1960-61. One of the interesting results brought out was that the performance on tests of students taking the televised course was not significantly different from that of students who took the course the year before under the same instructor in a large lecture hall and using conventional methods [4, p. 44].

Experimentation with closed-circuit television at Colgate University has been primarily designed to improve the quality of instruction. Beginning in 1964, in two of the six sections of the introductory course, "The American Economy" series of lectures was utilized in the first thirty minutes of each fifty-minute session. In the remainder of the period a panel of six students produced a live broadcast discussion. On the basis of the results of an examination common to all sections, it was concluded that the TV classes performed at least as well as the classes taught in the traditional manner.

In the spring of 1966 Colgate University tried a different approach, which they call "taped team teaching." Lester Blum reports that one of the three section meetings a week is devoted to a televised discussion by three staff members of some important topic of the course; or alternatively to an interview of a guest lecturer by one or two members of the staff.

At the University of Illinois, a one-semester, three-hour terminal course in the elements of economics has been given since 1959. The course has an enrollment of about 800 students, of whom from one-third to one-half are taught by television. D. W. Paden reports that "when checked from time to time against student achievement in 'conventional' sections of the course, the usual finding has been one of no significant difference."

The particularly interesting aspect of the Illinois course, however, is the way in which Paden and his associates have gone about successively improving the televised presentations. Their procedure was suggested to them by the success reported for programmed instruction in inducing the student to respond in accordance with carefully defined lesson objectives and in reinforcing his knowledge by inducing correct responses.[8]

Three lessons were selected for experimentation from about the middle of the course. Detailed objectives for each lecture were formulated in testable terms; test items were prepared for each lesson objective; individual students were taken through the instruction to determine how well the lessons realized the objectives; the individual lessons were then revised. The procedure was repeated several times. At each step, the programmed version of the lectures showed marked superiority over the preceding version.

In general, experience with closed-circuit television lectures seems to demonstrate that, if the choice has to be made between large lecture classes before which the lecturer appears in person and large classes

broken into smaller sections to hear and view a televised lecture, the latter is the preferable alternative. It is preferable, not primarily because still larger classes can thus be accommodated, but rather because the lectures can be better prepared, can utilize better visual aids of all descriptions, and can be pretested. Furthermore, the medium can be used to great advantage to supplement the teaching of smaller groups with, for example, student or faculty panel discussions, lectures by visiting experts, or short presentations of particular problems or timely issues.

REFERENCES

1. Rendigs Fels and D. R. Starleaf, "Controlled Experiments in Teaching Techniques," *So. Econ. J.*, July, 1963, pp. 68-73.
2. D. R. Fusfeld and Gregory Jump, "An Experiment with Programmed Instruction in Economics," *So. Econ. J.*, Jan., 1966, pp. 353-56.
3. M. L. Joseph, "Role Playing in Teaching Economics," *A.E.R.*, May, 1965, pp. 556-65.
4. A. H. Reede and R. K. Reede, *Televising Instruction in Elementary Economics*, Indus. Res. Bull. No. 5 (Bur. of Bus. Res., Pennsylvania State Univ., 1963).

THE EFFECTIVENESS OF PROGRAMMED LEARNING IN ELEMENTARY ECONOMICS*

By Keith Lumsden

Stanford University

Two years ago I read a paper, prepared by Richard Attiyeh and myself, which discussed the effectiveness of programmed learning in economics at the high school level.[1] Since then I have "graduated" to the college level, and if one indication of successful graduation is the ability to discover that unanswered questions are easier to find than are the answers, then this graduation has proved to be alarmingly successful.

As economists, one facet of our lives which consumes a considerable portion of our resources is teaching. Most of us know that while all of our colleagues are not the best teachers we, ourselves, are doing a "good job." But are we? We are often very concerned with the concept of efficiency in our learned journal articles, but are we seriously concerned with the concept of efficiency in the teaching of economics? It is encouraging to note that considerable resources are now being devoted to the whole question of economic education, and even if answers to interesting questions are not always forthcoming, at least some interesting questions are being asked.

This paper reports on several studies undertaken to measure the effectiveness of "programmed learning" in microeconomics at the college level. A programmed text is an exposition in which analysis proceeds by gradual degrees and in which continuous student participation in the form of written responses is required. Lessons are presented in small logical steps to allow the student to grasp simple concepts before proceeding to complex ideas and analysis. The material is broken into short paragraphs or frames, each frame containing at least one blank. The correct responses to the blanks appear at the bottom of the

[1] Richard Attiyeh and Keith Lumsden, "The Effectiveness of 'Programmed Learning' in Economics," *A.E.R.*, May, 1965. pp. 549-55.

page. While reading each frame the responses are covered and on the completion of the frame by the student he compares his written responses with the correct ones. The student can see immediately whether he is grasping the material; if his responses do not agree, he rereads the material until his understanding is clear. Since the text is a self-teaching device each student can proceed through the material at a pace most suitable to his individual capacity. This certainly permits a degree of flexibility which is absent, to a considerable degree, in the large lecture course.

The programmed material[2] used in all the studies I am going to describe was essentially the material contained in the text, *Microeconomics, A Programmed Book,* by Lumsden, Attiyeh, and Bach. This text analyzes the way a market economy functions and the conditions under which a market system will lead to an efficient allocation of resources. It is comparable to traditional texts in coverage of demand, productivity and costs, equilibrium of the firm, and price determination. It concentrates more on general equilibrium and economic efficiency and less on monopoly.

The first two experiments were similar in design. One occurred at Stanford (summer, 1965) and involved approximately 100 students with no previous college courses in economics; the other occurred at the University of California at San Diego (January, 1966) and involved approximately 130 sophomore students who had all taken one preceding course in macroeconomics. In each university the class was split into three matched groups.[3] In each experiment let me call the matched groups A, B, and C.

Students in group A were given rough draft copies of *Microeconomics* and were instructed to study only from this material in preparation for an exam to be given one month later.

Students in group B were taught conventionally by me. They had access to their texts (same texts as group C) and regularly attended lectures on the same material.

Students in group C were instructed to study only from material contained in two basic texts. (Bach and Samuelson at Stanford and Dorfman and Samuelson at U.C.S.D.)

All students were permitted to attend question sessions in which no lectures were given but in which questions were raised by the students, and all students were permitted to consult the professor and teaching assistants during office hours.

---

[2] Over time, there existed qualitative differences in this material which went from rough draft manuscript to the second printing of the text—the implicit assumption being that the text improved and the number of errors decreased.

[3] The matching at Stanford was based on pre instruction test scores on "The Test of Economic Understanding." At U.C.S.D., it was based on the grades achieved in the macroeconomics course.

At the end of one month all students took the same examination, which was a combination of true-false and essay questions.

In both experiments the results were clear and statistically significant. Students in group A, the programmed section, outscored the students in the other sections when the complete examination was considered. There was no statistically significant difference, however, between any two groups on the "objective" part of the examination.

After the examination at Stanford, all students met regularly in class for the remainder of the course and were taught conventionally. When I gave this class the final examination covering predominantly macroeconomics, I compared the final examination results by sections

TABLE 1

TEST RESULTS

	Number of Students	Average Examination Scores		
		Total Score	True-False	Essay
Group A (programmed text)................	44	47.8	8.8	39.0
Group B (lectures+ texts)...............	44	34.7	8.2	26.5
Group C (texts only)....	43	32.0	5.8	26.3

A, B, and C and discovered that those students who had been in section A, the programmed section, again outscored the other two sections. Unwisely, I threw out the data, concluding that, despite my selection method, I had a biased sample; i.e., the students in the programmed section were "superior" to the students in the other sections. Now I am not sure. It may well be that the programming approach interested students significantly, gave them a better basis for continued study, or changed their study habits for the better. Because I threw out the data I do not have "numbers" from the Stanford experiment but I can present the "numbers" from the U.C.S.D. experiment.

A rank order test[4] was performed pairwise on the three groups and the results substantiate that the scores in group A (the programmed section) are greater than the scores in either other section. This was true for total scores and for essay scores. There was no statistically significant difference in the total true-false scores between any two groups nor between any of the scores for groups B and C. The results are summarized in Table 2.

[4] H. B. Mann and D. R. Whitney, "On a Test of Whether One of Two Random Variables Is Stochastically Larger Than the Other," *Annals of Math. Statis.*, Vol. 18, 1947, p. 50.

TABLE 2

VALUES OF TEST STATISTIC U

	A vs. B	A vs. C	B vs. C
Total score.................	2.45	3.00	0.10
Essay score.................	2.69	2.87	−0.49
Total true-false score.......	−0.93	0.74	0.68

A parallel but slightly different experiment was carried out in the fall quarter 1966-67 at four different institutions, in each case using similar procedures. Sections were randomly selected from elementary economics classes and, in the part of the course covering the core of micro-theory, were instructed to study only from *Microeconomics*, the programmed text. The remainder of each class was taught in the conventional fashion without access to *Microeconomics*. The material covered in class paralleled the contents of *Microeconomics*. At the end of four or five weeks all students were given the same examination, which was a collection of multiple choice questions (each having four possible answers) taken primarily from workbooks accompanying the major elementary texts. There were forty questions; no penalties were given for wrong answers; the test scores are summarized in Table 3.

Table 4 is a summary of a simple regression analysis in which test score (Y) is the dependent variable and intellectual capacity (X_1) and programmed-nonprogrammed (X_2) the independent variables. Unfortunately because of diversification among schools in measures of intellectual capacity it was not possible to pool results. Programmed-nonprogrammed is a dummy variable with values 1 and 0. Thus the equation reads

$$Y = \alpha + B_1 X_1 + B_2 X_2$$

TABLE 3

TEST SCORES

Institution	Number of Students		Textbook Used	Average Score	
	Pro- grammed Section	Regular Section		Pro- grammed Section	Regular Section
Marin Junior College..	30	37	McConnell	20.4	23.7
University of Illinois...	32	287	Samuelson	21.7	21.7
Wesleyan University...	54	76	Lipsey & Steiner and Samuelson	25.4	24.6
Yale University.......	8	14	Samuelson	28.5	24.9

B_1 we would expect to be positive and B_2 positive if programming added to the score and negative if the reverse were true. Data on intellectual capacity of students was unavailable for Illinois.

The only B_2 value which was statistically significant was for Marin College. Students at Marin, studying solely from the programmed text, did not do as well as students conventionally taught; at the other three colleges there was no statistically significant difference between the scores of the programmed and conventionally taught sections. Now a devotee of programmed learning could argue that Marin was a bad example. Perhaps junior college students do not study conscientiously on their own. This hypothesis is supported when time spent studying is

TABLE 4

REGRESSIONS AND AVERAGE SCORE

Institution	Regression Equation	Average Score (Y)		
		Combined	Pro-grammed	Non-programmed
Marin........	$Y_M = 15.64 + .009X_1 - 3.24X_2$	22.2	20.4	23.7
Illinois.......	$Y_I = 21.67 \qquad + .00X_2$	21.7	21.7	21.7
Wesleyan.....	$Y_W = 2.00 + .028X_1 + .73X_2$	24.9	25.4	24.6
Yale.........	$Y_Y = 15.77 + .114X_1 + 3.54X_2$	26.2	28.5	24.9

included as an independent variable. Then there is no statistically significant difference between the scores of the programmed and nonprogrammed sections. Perhaps those conventionally taught had an outstanding instructor—the relatively high scores about the same as for the four-year colleges for regular sections would certainly support such an hypothesis. In the other instances, the B_2 coefficients were positive though not significant.

I believe the results to be important for the profession, for students, and for efficiency in teaching. Accepting the examination given and the procedures followed, teaching the core of price theory solely by the programmed text saves both teachers' and students' time and is at least as effective as regular classroom instruction.

The results are most impressive for an institution such as Wesleyan generally recognized for the high caliber of instruction. A further check was carried out at Wesleyan. The students were subjected to an additional examination, problems plus essay, made up and graded by a professor who was not otherwise involved in the experiment. Again the results were consistent; students in the programmed section performed as well as those conventionally taught.

If we return to the Stanford and U.C.S.D. experiments reported

above, you will recall there was no significant difference in the objective examination results between programmed-nonprogrammed groups but in the essay examinations the programmed sections significantly outscored the nonprogrammed sections. In both cases the essay question was broad; students were asked to appraise the efficiency of the price mechanism in a market economy. At Stanford the grading was done by a teaching assistant. At U.C.S.D. the grading was done by teaching assistants and secondly as a check by a Stanford faculty member. While there is some disagreement between the graders on a few papers, the overall results remained unchanged. One possible explanation, which modesty almost prohibits me from offering, is that the programmed text, because of its content, prepares students for an examination question of the type described better than do traditional texts plus instructors. I think a more reasonable explanation is that students obtain through the programming technique a better understanding of the basic tools and in addition have been using those tools in writing their responses and are then better equipped to use those tools in a broad essay question. Since it is almost practically impossible to administer essay examinations on a national scale, my current plea is for more testing of this nature by you in your individual colleges and universities.

Let me now return to the question of efficiency in the teaching of economics. If one accepts the hypothesis that certain portions of the principles course can be just as effectively taught by programmed texts as by regular instruction,[5] then scarce professorial time can be used to teach other portions of the subject, introduce more material, meaningfully supplement existing material, or give more individual attention to students. (The professor could, of course, also have more leisure, or if returns were still positive, devote more time to research.)

In the efficiency context I do not intend to discuss optimal allocation of student time. Table 5, however, summarizes the average time spent by students in programmed versus nonprogrammed sections at the various institutions.

One striking fact is the substantially less time spent by students in the programmed sections, though they apparently learned as much as students conventionally taught. Not only was the average number of hours smaller for the programmed sections, but the students in those sections could choose which hours they devoted to studying microeconomics. While such a statement is also applicable to the hours used for studying outside of class by conventionally taught students, relatively

[5] For an interesting experiment in the use of television, see Donald Paden, "Television as an Instrument of Instruction" in *New Developments in the Teaching of Economics* (Prentice-Hall, at press).

TABLE 5

STUDY TIME SPENT

	Average Hours Spent Studying	
	Programmed Section	Nonprogrammed Section
Marin Junior College.........	16.0	27.3
University of Illinois..........	9.5	19.5
Wesleyan University.........	22.0	27.0
Yale University.............	Section had text for eight days prior to the examination	Section was taught for six weeks prior to the examination

few institutions give students a wide choice of the timing of hours of instruction.

In addition to being asked how much time they spent studying, students were also asked whether they preferred programmed instruction to conventional instruction, were indifferent, or preferred conventional instruction on the basis of this study. The results, which are mixed, are presented in Table 6.

The Illinois and Wesleyan students certainly appear to appreciate their instructors more than other students. In analyzing responses, the main reasons given for preferring programmed instruction were ease of learning, being forced to learn, avoidance of learning wrong things well, and flexibility of study time. The main reasons for preferring conventional teaching were ability to discuss issues in class with the instructor and learning from the questions of other students; the bulk of the Wesleyan answers were of this variety.

My simple summary of my experiments would be that we do not have all the answers. In fact, we have relatively few answers about

TABLE 6

PREFERENCES

	Percentage		
	Preferred Programming	Indifferent	Preferred Conventional
Marin Junior College............	68	23	9
University of Illinois.............	21	29	50
Wesleyan University.............	40	18	42
Yale University.................	50	38	12
Stanford University[6]............	67	17	16

[6] The Stanford figures are for this year's entering students in the Graduate School of Business taking a course in microeconomics and business applications. Many of those students had prior training in economics as undergraduates. They were required to study the text *Microeconomics* in the first week of the quarter, after which they were given an examination.

programmed instruction in economics. My plea is for more experimentation. To textbook authors and devoted teachers on the basis of these experiments, my presumption is that programmed instruction will replace neither but appears to be more efficient than either for parts of the principles course.

Perhaps the best summary of my position is that stated by a noted psychologist:[7]

Question—Do students learn from programmed instruction?
Answer —Clearly, yes.
Question—Is programmed instruction more efficient than other forms now in use?
Answer —Improper question. It is more efficient than some, not more efficient than others. Please specify type of instruction, level of students, quality of instructor, nature of content, and method of evaluation to be employed.

[7] Professor A. Bavelas in an unpublished paper (Stanford, 1966).

A NEW "TEST OF UNDERSTANDING IN COLLEGE ECONOMICS"*

By Rendigs Fels

Vanderbilt University

On recommendation of the Committee on Economic Education of the American Economic Association, the Joint Council on Economic Education has commissioned preparation of a new Test of Understanding in College Economics. Three considerations motivated the decision.

First was the success of the existing high school Test of Economic Understanding (TEU), which has been taken by hundreds of thousands of high school students and yielded much of the data used in two interesting papers by Bach and Saunders published in the *American Economic Review* [1] [2]. Although the TEU has often been used for college students, it does not test much of the knowledge that professors try to impart in the sophomore course. Most of the TEU questions are on what might be called layman's economics—the kind of knowledge that intelligent people can pick up without systematic effort —as distinct from economists' economics, which requires sustained study. The fact that the TEU gets used for purposes for which it was not designed showed the need for a comparable instrument for evaluating what college students learn.

Second was the response to the call by the Committee on Economic Education for reports on experiments in teaching basic economics. These experiments, which Bernard Haley has discussed in a previous paper at this session, were with one exception lacking in evaluation. The experimenters, though, could hardly be blamed. There was not available any generally accepted test for measuring the results of experiments.

Third was the hypothesis advanced several years ago by George Stigler at the meetings of this Association that the elementary college course in economics has no lasting value. Stigler proposed taking a sample of seniors, or better yet, men five years out of college, "equally divided between those who have never had a course in economics and those who have had a conventional one-year course. Give them an examination on current economic problems, not on textbook questions. I

predict they will not differ in their performance" [4, p. 657]. The data of Bach and Saunders, based on the TEU, tend to confirm Stigler's hypothesis [2], but a better instrument for testing it was needed.

All this does not imply that no good tests have been previously available. Simon Whitney has done pioneering work in measuring the success of the elementary course [5][6]. Though handicapped by slender resources, which did not permit going beyond the true-false type of question or employment of measurement experts or use of an appropriate sample to compute national norms or even preparation of test specifications, he achieved interesting results.

The Educational Testing Service has a more elaborate test than Whitney's, available at a charge of $5.00 per student, which is designed to evaluate students applying for advanced standing or credit. The use of this test for research purposes is precluded by tight security regulations which would prevent publication of detailed research findings. Other good tests that have come to my attention are also not available for research.

As a result of the Economic Education Committee's recommendation, made a little over a year ago, the Joint Council undertook to provide a substantial amount of money, entered into an agreement with a testing organization (the Psychological Corporation) for its professional services, and appointed a committee to supervise construction of the test consisting of six economists (G. L. Bach, William G. Bowen, R. A. Gordon, Paul A. Samuelson, George J. Stigler, and myself) and an outstanding authority on psychometrics, Paul L. Dressel. John R. Stalnaker, President of the National Merit Scholarship Corporation and chairman of the committee that developed the TEU, is serving as consultant, and a number of other economists are cooperating by writing questions for the test.

The Committee decided on the multiple-choice form of question. The test will be in two parts, each designed for administration in one class hour. Part I will be on the content of the typical first semester of college economics, principally macroeconomics. Part II will be on the content of the typical second semester of the elementary course and will emphasize microeconomics. For those professors wanting a one-hour test on the work of both semesters, there will be a hybrid consisting of half the questions from each part. There will be matched alternate versions of each part of the test, making it possible to give the class one version at the beginning of the course and the other at the end to measure the progress of the students. These arrangements will permit the user of the test a great deal of flexibility.

Specifications, pretesting, and norming are important steps in test construction. The specifications were prepared last summer. They

consist of a detailed outline of the content to be covered by the test with a statement of the percent of questions to be devoted to each of the main headings and subheadings. In Part I, 20 percent of the questions will concern scarcity, the functions of all economic systems, and some simple aspects of supply and demand; 8 percent will be on macroeconomic accounting; 16 percent on income-expenditure theory; 18 percent on money, banking, and monetary policy; 16 percent on macro aspects of fiscal policy; 8 percent on determinants of economic growth; and 14 percent on stabilization and growth policies. In Part II, 20 percent of the questions will be on competitive markets, including agriculture; 30 percent on the theory of the firm and antimonopoly policy; 15 percent on factor markets and income distribution; 10 percent on the government's role in the allocation of resources; 15 percent on international economics; and 10 percent on comparative economic systems.

For each part as a whole, the specifications call for equal numbers of questions in three categories: recognition and understanding, simple application, and complex application—terms which will be explained below. The specifications imply that a substantial fraction of the questions will involve policy issues—a larger fraction than the percentage allocation of questions by subject might suggest. Understanding of nonpolicy subjects can often be best tested in a policy context. Most important, we are making a concerted effort to get questions that are fresh and imaginative.

The purpose of the specifications is to ensure that the test will measure what the Committee thinks it should measure, giving due weight to the most important concepts, instead of consisting of a haphazard collection of questions. The specifications also give due warning to professors whose teaching objectives differ, so that they can interpret results of the test correctly.

Pretesting means trying out questions on a large number of students to help decide which questions will go into the test. Pretesting for Part I will take place in January, 1967, for Part II in May. We shall pretest about twice as many questions as we expect to use. Each question needs to be tried out on 700 to 1,000 students; each student taking part will be given about 30 questions; over 200 questions will be pretested for each part; so that altogether 7,000 students will be involved in pretesting for each part. Glenn Burress, of the Joint Council, who is in charge of arranging for pretesting, has had excellent cooperation from the schools he has contacted. We are grateful for the generous help.

One purpose of pretesting is to eliminate questions with bugs. Writing multiple-choice questions that are airtight is astonishingly difficult.

Even the careful review and editing that all questions are given before pretesting cannot be expected to detect all the flaws that can mislead sophomores. To illustrate, one question, which we have already thrown out, involved the relation of the terms "investment" and "capital formation." The "right" answer, according to the author of the question, was supposed to be that they are nearly synonymous, as in fact they are in the national income accounts. Several other economists agreed that that was the right answer. Fortunately, we gave the question to a small number of students on an examination. In trying to find out why most students got such a simple question wrong, we came to realize that another one of the choices could be considered right; namely, that capital formation is a subheading under or illustrative of investment. That choice could be correct, since economists sometimes use "investment" broadly to include investment in human beings. But this discovery did nothing to unravel the mystery of why most students gave the wrong answer, for not a single student had chosen that particular option. The majority had answered that investment was a precondition for or necessary to capital formation. And when we thought about it, we realized that they could not be called wrong. They were thinking of investment in the dictionary sense of "the investing of money or capital in some species of property for income or profit."

Pretesting does more than eliminate faulty questions. It reveals which questions have the power to discriminate. Questions that everybody gets right or that everybody gets wrong are not very helpful. Generally we shall seek questions that are answered correctly by 30 percent to 70 percent of the students. Nor are questions helpful if poor students get them right more often than good students. Correlation of right answers with some measure of the quality of the student will be determined. The rule of thumb is that the correlation coefficient must be at least + 0.25 or + 0.30.

In addition to eliminating faulty questions and those that do not discriminate, pretesting will provide preliminary norms. Norming means the development of standards of performance defined, usually, by the attaining of various scores on the test. Norms are needed so that research workers and experimenters can compare results in their samples with a good cross-section of students across the nation. The temporary use of preliminary norms will make it possible to publish Part I of the test in September, 1967, and Part II in January, 1968. Revised norms will become available in the following year. A manual will be published describing in detail the way the test was developed and the uses that can be made of it.

What can the new test accomplish? Let me take up the negative side first. Its purpose is not—repeat, not—to evaluate the skill of individual

teachers or help deans decide on raises and promotions. Aside from the fact that may professors would not use the test if they thought the results might be used against them, the Test of Understanding in College Economics will measure accomplishment of only one of the objectives of the elementary course. It does not purport to measure the extent to which a professor broadens the mental horizons of his students or stimulates their intellectual curiosity or inspires them to noble endeavors. It will not necessarily even be a good gauge of how well prepared students are for a course in intermediate theory.

Nor will the new test provide a ready-made final examination for weary professors who do not want to write their own. You might get away with using the test for grading purposes once, but even that is doubtful. Although the Psychological Corporation, which will sell the test at a small charge, assures me that it will not fill orders from fraternity houses, tests available by direct purchase, regardless of precautions, ultimately become available to anybody.

Rather than evaluating individual teachers or individual students, the new test will be a research instrument. Lumsden's paper at this session has described an example of the kind of research we have in mind [3]. Unfortunately, we were not far enough along in the development of our test to be of much help to Lumsden and the others associated with that project. We have, I hope, been somewhat more helpful to Campbell R. McConnell, who is conducting a controlled experiment on the comparative effectiveness of closed-circuit television versus live teaching of elementary economics. We have made our file of questions available to McConnell for testing the students receiving the two different kinds of instruction. Our finished test will, we hope, be considerably more valuable for research of this kind.

The test will measure the extent to which groups of students gain in understanding of college economics. In this context, economic understanding does not mean the ability to manipulate abstract models or to draw complex diagrams with the curves labeled correctly and sloping in the right direction. There will be a few diagrammatic questions, but only the simplest sort. Instead, the test will emphasize the ability to apply economic principles to real problems, including issues of public policy. More specifically, the plans call for equal numbers of three kinds of questions: (1) "recognition and understanding," (2) "simple application," and (3) "complex application."

"Recognition and understanding" refers in principle to questions that could be answered out of a textbook. Here is an example that, in a slightly different form in a preliminary tryout, turned out to be rather difficult, more difficult in fact than the example of a simple application to be given later:

Increased aggregate spending, such as can be generated by expansionary monetary-fiscal policy, has limited effectiveness in speeding the growth rate of the U.S. for extended periods such as several decades because

*1. The growth rate depends mainly on nonmonetary factors.
 2. Investment is relatively insensitive to low interest rates.
 3. The U.S. is a high-saving economy.
 4. Population growth is the basic determinant of long-term growth.

The answer to this question is stated clearly, though in different words in the Task Force Report [7, p. 53]. Since it presumably is part of what most of us teach, we classify this question as "recognition and understanding." Items in this category need not and should not be answerable by rote memory. The better items of this type involve restatement or recognition of an idea in somewhat different language from that in which it was originally learned. They may call for explanation, summarization, or simple extension of an idea. Thus such items can and should test understanding or comprehension rather than recall. We need such questions in the test to distinguish between those students who know economic principles but cannot apply them from those who do not understand the principles at all.

A simple application of a principle or concept may be defined as one in which the student demonstrates that he can utilize it when its use is specified or clearly implied. A complex application may require that a student demonstrate ability to select and utilize a concept or principle when its use is not specified, or it may require that two or more concepts or principles be related in some manner not previously presented to the students.

Here is an example of simple application:

In public discussion about the current inflation, Administration economists have failed to mention the "multiplier effect," the theory that a billion dollars of deficit-spending by the federal government finally creates $2½ to $3 billion of consumer purchases. But if this theory is true, it certainly must work in reverse. *The net effect of a billion-dollar cut in federal spending today would be the equivalent of a 3 billion-dollar increase in taxes.*

The economic analysis in the underlined sentence is:

 1. Essentially correct.
*2. False because an increase in taxes is subject to a multiplier effect, too.
 3. False because taxes have a multiplied effect but government spending does not.
 4. False because neither taxes nor government spending has a multiplied effect.

This question requires the student to understand the multiplier well enough to evaluate a quotation. The quotation was not invented for the test but was adapted from a nationally syndicated newspaper column. One of the best tests of economic understanding is whether students can detect nonsense when they see it. Regrettably, the opportunities to see it are abundant.

Now for an example of complex application:

To raise the prices of farm products and the incomes of farmers, the government purchases crops and stockpiles them. An alternative would be outright money income subsidies that

are not related to farmers' output. Both policies increase farmers' demand for industrial products. Both use productive resources. What can be said about the net effects of the two alternatives on the real income of the rest of the nation in times of full employment?

1. Both reduce it by about the same amount.
2. Both increase it by about the same amount.
3. Both increase it, but subsidies increase it more.
*4. Both reduce it, but stockpiling reduces it more.

In this question, the student must analyze and compare two alternatives, making it a complex application.

The three examples just cited all happen to involve policy questions. The Committee has not yet decided how many policy questions to include. Probably they will be a minority. But two-thirds of the questions will require application. In this respect, the Test of Understanding in College Economics will be a departure from most of the economics tests I have seen.

In the mixed economy that now prevails in the United States, Congress every year enacts new legislation regulating economic life. Up till now, this kind of economic system, which is of rather recent date, has performed with great success. Whether it continues to do so or becomes bogged down in a maze of ill-considered rules will depend on how well the American public can apply economic principles to solving its problems. The Test of Understanding in College Economics is designed to measure how much the principles course contributes to America's economic understanding.

REFERENCES

1. G. L. Bach and Phillip Saunders, "Economic Education: Aspirations and Achievements," *A.E.R.*, June, 1965, pp. 329-56.
2. ———, "Lasting Effects of Economics Courses at Different Types of Institutions," *A.E.R.*, June, 1966, pp. 505-11.
3. Keith Lumsden, "The Effectiveness of Programmed Learning in Elementary Economics," *A.E.R.*, May, 1967.
4. George J. Stigler, "Elementary Economic Education," *A.E.R.*, May, 1963, pp. 657-59.
5. Simon N. Whitney, "Measuring the Success of the Elementary Course," *A.E.R.*, Mar., 1960, pp. 159-69.
6. ———, "Tests of the Success of the Principles Course," *A.E.R.*, May, 1965, pp. 566-71.
7. *Economic Education in the Schools*, Report of the National Task Force on Economic Education (New York: Committee for Economic Development, 1961).

DISCUSSION

WILLIAM R. ALLEN: Professor Fels provides a straightforward report on the preparation of a new college level test of economic understanding. Doubtlessly, the task of preparation is being performed systematically—and with expenditure of "substantial" resources—and there is little point in quarreling over most of the conspicuous details of the planned test. But there are some matters, perhaps of a less obvious nature, which raise questions and qualms.

Professor Fels finds a "need" for an instrument to evaluate "what college students learn" and to measure the results of experiments in teaching basic economics. Who is to use the test and why? And can one test of acceptable length evaluate all experiments? One might suppose that a competent teacher could do his own testing and that one not sufficiently clever and skilled to do so would be quite incapable of adequately adjusting his course if the grades of his students in the uniform test convinced him that adjustment was in order.

It is indicated also that the new test is required in order to test Stigler's hypothesis that college students and graduates who have not taken a principles course would do as well in "an examination on current economic problems" as those who have lost this academic virginity. Preparing an elaborate test for this purpose is, to quote Adam Smith out of context, quite an unnecessary attention, for the validity of the hypothesis is such a sure thing as to smack of an axiom. The typical student who has taken a typical economics course from a typical instructor using a typical textbook—and I doubt that general use of the teaching experiments described by Professors Haley and Lumsden would greatly alter the picture—is in a position similar to that of the mathematics student who knows the rules of differentiation but does not comprehend the meaning of dy/dx. And if Stigler's hypothesis is confirmed, is the profession then obligated to reshape its texts and courses until it is refuted?

Next, consider the format and structure of the test. The questions are of the multiple-choice type. This or a similar type of question presumably is dictated by the physical problems of grading—but it would seem also to preclude discernment of gradation of sophistication in the answers. Most good questions for sophomores are appropriate for Ph.D. qualifying examinations, and many good Ph.D. questions can reasonably be put to sophomores—but one hopes that the quality of the answers would distinguish the one group from the other.

This multiple-choice test is divided into two parts: the first devoted mainly to aggregative economics and the second to price and allocation analysis, with questions of three kinds—recognition and understanding, simple application, and complex application. Economic theory is essentially price theory: the theory of resource use and exchange, of allocation and absorption, encompassing the incentives to act, the criteria by means of which actions are judged and plans are formulated, and the manner in which decisions are made and enforced, dealing with alternatives and their costs and the ways of learning and communicating the preferences, capabilities, offers, and options of both con-

suming and producing agents. Slowly and belatedly, bridges are being con-structed from price theory to aggregative theory, making the latter a consis-tent and integrated portion of general economic theory, but there appears to be little analytic feedback from aggregative theory to price theory. Some of us, therefore, are unhappy to see the natural sequence of coverage reversed. Even those who have no qualms about multiple-choice questions and who ap-plaud putting first things second may still be displeased by individual ques-tions. For example, recognition and understanding questions are those that "could be answered out of a textbook." Every textbook? Or are we to banish those which do not provide quite explicit answers to those questions? With re-spect to the questions generally, there can be difficulty in attaining unanimity on what constitutes truth. Professor Fels kindly provided me with a copy of the "specifications" for preparation of the test. These specifications include parenthetical comments of definition and supposed fact on many topics in order to orient and guide those preparing questions. One might have thought that this, too, is a case of an unnecessary attention—but, discouragingly, I find myself in some measure of either disagreement or wonderment with a con-siderable number of these supposedly elemental comments. Along with doubts concerning what is included, I fret about what is omitted. In particular, I find no price theory and essentially no monetary theory in the national income sec-tion; price theory is overwhelmingly concentrated on the theory of the firm and of market structures, with, for example, the principle of comparative advan-tage relegated to a late section on international economics; and the importance of time—in the calculation of cost and profit and in speculation—appears to be wholly missing.

In what is omitted, as well as in the sequence of topics, the test probably reflects the choices and emphases of most of the profession. And that interests me a good deal more than how well sophomores do in a common multiple-choice test. I am less concerned now by how well the lessons are being learned than I am about what lessons are being taught. Indeed, considering the quali-ty of what is being taught, it may be best that not much be learned.

According to Professor Fels, the test is "designed to measure the extent to which groups of students gain in economic understanding"; it "will emphasize the ability to apply economic principles to real problems"; "a substantial fraction of the questions will involve policy issues"; and "two-thirds of the questions will require application." But how is it to be determined what con-stitutes "economic understanding"? If the policy questions are to go beyond mere mechanics, how is it to be determined what are the correct answers? If applications are to be the bulk of the test, how is it to be determined what are the included areas of application? Often enough it is difficult to determine truth on even formal, technical points, but on matters of orientation, per-spective, emphases, and philosophical criteria, are we to follow (and how is the choice to be made?) Friedman or Galbraith, Stigler or Samuelson, Bucha-nan or Tobin, Haberler or Heller, Hazlitt or Keyserling?

The test—while, curiously enough, not "necessarily" providing "a good gauge of how well prepared students are for a course in intermediate theory" —"is designed to measure how much the principles course contributes to

America's economic understanding." And America's future—in contrast to her past—"will depend on how well the . . . public can apply economic principles to solving its problems." (In meetings of other associations, it is probably being claimed that America's future depends on how well the public understands principles of history or political science or sociology or whatever.) I am compelled to hope that Professor Fels is wrong, for, if our future depends on the economics principles course, we might well be excused in relaxing and trying to enjoy the inevitable disaster.

DONALD W. PADEN: The first two papers presented here attest to the experimentation which is going on in the teaching of the first course in economics. The third paper describes an attempt to provide the teaching profession with an evaluation instrument which should make future research on the effectiveness of such experiments considerably easier than has been true in the past. With respect to the report by Professor Haley, one would like to have available the details of the various projects to assess the quality of the research effort involved and the statistical sophistication with which it was carried on. Experimentation, although encouraging, is not enough. The researcher must have in mind where he is going and some generally acceptable way of demonstrating whether he has arrived at his goal. The generation of data to support clearly stated hypotheses with respect to student learning is an undertaking beset with great difficulties and is perhaps generally underrated by members of our profession.

The Lumsden paper, which I have been asked to discuss specifically, presents at least two aspects which deserve comment. For the orientation of the uninitiated, the programmed text is properly described in the study as one which incorporates logical sequencing, student participation (usually involving written responses), reinforcement (generally through immediate knowledge of the correct answer), and a measure of self-pacing. Among professional programmers at the present time there is a movement to reformulate programming, not in terms of the end result (e.g., the programmed text), but as a process through which the product described above is produced. This idealized process involves (1) the specification of detailed objectives covering the subject to be taught stated in behavioral terms, (2) the concurrent development and refinement of testing instruments to be used to measure whether the objectives have been achieved, (3) the production of materials (as described by Professor Lumsden), and (4) the generation of data for a specified population (or populations) indicating the proportion of students in the population who, after using the materials, exhibit the competencies stated in the objectives—as measured by the testing instrument.

This process, of course, goes far beyond the production of materials incorporating the programming format. In addition to what is reported by Professor Lumsden, it suggests the desirability of carefully stated objectives, a fully validated testing instrument and data on the proportion of students who, after the instruction, exhibit the behavior thought desirable by the programmer (other than what can be inferred from overall examination scores). With respect to the latter point and as one of a very large number of examples, it

would be illuminating to learn that, after exposure to the Lumsden program dealing with demand elasticity, 90 percent of the students at Stanford University recognized that a decrease in price will be accompanied by an increase in total revenue given an elastic demand schedule (as compared with, say, only 50 percent of the students who use a textbook). Without being critical of the Lumsden research, this is the kind of information which the prospective user needs. What is reported in the paper is interesting and encouraging; what is omitted is also vital—at least before clear-cut decisions can be reached as to what tasks can be accomplished most effectively through programming and what tasks should be tackled in other ways.

Turning to the actual experiments reported upon in the paper, it is highly encouraging that Professor Lumsden was able to involve different institutions, that different textbooks were used, that a variety of instructional methods were represented, and that the results—evaluated by a common evaluation instrument—seemed to indicate that students did as well with programming as when taught by conventional methods. Even discounting somewhat for the natural enthusiasm of the investigator, there seems to be an undertone of acceptance both among the institutions and the students involved. Given the independent nature of the teaching profession, the mere fact of interinstitutional cooperation itself should be regarded as a major accomplishment.

The claims made by Professor Lumsden in his summary statement are extremely modest. What follows, therefore, is a warning to the reader not to jump to conclusions which Professor Lumsden sought to avoid. Indeed, it is hoped that for the benefit of prospective users of the program, future work will provide the sort of information which has been suggested above.

First, it should be pointed out that the study reports upon the effectiveness of a particular programmed text in economics rather than, as the title might suggest, the effectiveness of programmed learning in any general sense. Indeed, the latter is perhaps an unattainable objective given the fact that teaching involves particular teachers each with unique competencies, that a particular program must be used in each experiement, and that a particular evaluation instrument must be used with unknown biases. Even though the results had been negative, the confirmed programmer would have regarded this as but one of many possible experimental programs—in the same spirit as the medical researcher would not regard one failure in the treatment of a disease as grounds for giving up his research.

Second, the study suffers, somewhat, from the fact that probability samples were not used in all of the schools involved. For example, at the University of Illinois one section is reported to have been selected "at random." This is far different from selecting students at random and raises questions as to the propriety of reporting for this school "no statistically significant difference between the scores of the programmed and nonprogrammed sections." This observation is not appropriate in the case of the University of California at San Diego, where matched groups were used, as was true at Stanford University but for which data are lacking.

Finally, there are a number of small but troublesome problems which were not fully reported. (Perhaps within the confines of the paper, they could not

be covered adequately.) What precautions were used to insure that the testing instrument did not favor one or the other type of instruction? Were the essay examinations graded independently in their entirety by different graders and are the scores available for inspection? Does the programmed material really cover the subject matter with the same intensity as the various textbooks? What were the measures of variability of the means so that the reader may judge for himself the statistical significance of the performance reported? Is there assurance that students who were instructed to use only the programmed material did not in fact use both the programmed material and the textbook (and possibly attended lectures as well)? And so on.

Certainly, the most striking fact reported in the study is the saving of time by students using the programmed materials. The average reported time spent in studying of roughly fifteen hours (an average for the three schools) covering a period of four or five weeks seems on the small side, even for students using programmed materials. Given the pressures on students for grades, the temptation to study the textbook might be overwhelming. Until assured that adequate safeguards were in fact employed, the reader must interpret the data on "hours studied" with caution. What the data do suggest is the need for further study with tighter controls.

As a matter of general information for economists interested in programming, there is an elaborate Joint Committee Report by the American Education Research Association, the American Psychological Association, and the Department of Audio-Visual Instruction on "Recommendations for Reporting the Effectiveness of Programmed-Instruction Materials." The general nature of the recommendations of this group is as follows: "The basic premise of this report is that instructional effectiveness must be judged for each program according to its demonstrated merit. Evidence for the effectiveness of a program should be based on a detailed study which has been fully documented in a technical report."

Until further experiments are conducted whose statistical design is more sophisticated than seems to be the case of the present study and until such a documented report is prepared, the substitution of the specific program reported upon here either for conventional lectures and/or for conventional texts must be based as much upon intuition as upon the research findings themselves. Although the present study is indeed convincing, the questions which have just been raised are testimony to the enormous practical difficulties in attempting to demonstrate, statistically, the superiority of one "method" of instruction over another. Using as a foundation what has been reported upon here, it is hoped that the present study will stimulate further research in programmed learning.

Professor Lumsden's contribution lies in making programmed materials in economics vastly more credible as a means of instruction than was the case before his research. The profession is correspondingly in his debt.

PHILLIP SAUNDERS: Professor Haley has done a commendable job of summarizing and reviewing a very unwieldy body of material. In reviewing his conclusions on "experiments" with the content of introductory economics

courses, I am struck by two points. The first is the general lack of systematic attempts at controls and evaluation in these "experiments." Second, despite the fact that "quite a few institutions are experimenting with different types of introductory courses for meeting the needs of different types of student," no mention is made of one type of student that I feel needs a great deal more attention—the future teacher.

In the area of the different techniques being employed to improve or supplement the large lecture sessions, the lack of evaluation is not quite as bad as it is in other areas. But, with one or two notable exceptions, it is still far from adequate. And, while the University of Virginia's use of television indicates the continuing attention that is being paid to the problem of the economics training of in-service teachers, in-service programs at this time are often faced with the types of problems associated with putting spilled milk back into bottles or locking barn doors after the animals have departed. More attention to the economics training of future teachers before they graduate from college is not only a more efficient way of getting at the basic problem, but it would also permit us to use the resources devoted to in-service training in a more imaginative and sophisticated manner than is possible in current programs which have to start from scratch.

The general lack of systematic controls and evaluation in the areas of course content and teaching techniques is disturbing in a profession that prides itself on the precision and rigor of the analytical tools increasingly being brought to bear in various areas of its research work. No one would think of publishing a research paper or making a public policy recommendation based on the kinds of evidence and evaluations we daily accept in making curriculum decisions. Yet, in the areas of teaching and course content, too many of us seem quite willing to "fly by the seat of the pants" in a manner that would be subject to well-deserved ridicule if we attempted to use it in other areas of our professional activity.

Professor Fels has suggested that one reason for the lack of evaluation in teaching experiments up to this point has been the lack of "any generally accepted test for measuring the results of experiments." This may be so. If it is, the work of Professor Fels and his committee seems to be going a long way toward making exciting breakthroughs in this area. But we must also remember that, as Professor Fels carefully points out, the new Test of Understanding in College Economics is designed to measure the comprehension and application of certain carefully specified content only.

It is quite likely that many of our introductory courses have other objectives in addition to specific content comprehension and application. Increased interest in economics, for example, is a possible course objective that might not be adequately measured unless we make specific attempts to measure interest as a separate variable in itself. Things like student reading habits and course selection after the introductory course are examples of some kinds of evidence that we might consider in addition to test scores and questionnaire responses in evaluating particular experiments. The fact that we have not made much use of this kind of evidence in evaluating past experiments proba-

bly means that we should take special care to consider several different types of evidence once a generally accepted test of content comprehension and application does become available.

Do special, more mathematical sections of introductory economics for engineers make them more interested in taking additional courses in economics than regular sections? Does emphasis on analyzing current policy issues increase students' ability to analyze different policy problems five and ten years after they are out of college? Do courses in economics affect students' social attitudes and value judgments as they, hopefully, learn more about the way in which our economy operates? There are many interesting questions which can and should be investigated but which the new Test of Understanding in College Economics, if used alone, will be able to answer only partially, if at all.

Professor Lumsden's paper gives an example of how evidence other than test scores can be used in evaluating experiments. And, in the area of testing, his methods indicate that in some circumstances it is not impossible to use essay questions as evaluation tools. To really test the merits of different teaching techniques such as programmed instruction or games and simulation exercises, however, I feel that we have to follow up and test several semesters (or several years, if possible) after the completion of the introductory course. Even though there have been several experiments that indicate that students in small discussion sessions do not score significantly better than students in large lecture sessions on final exams, temporary acquisition of knowledge is only one part of the learning process. If we also consider retention and the ability to transfer knowledge to new or different situations, I feel that the techniques that require more active student participation and involvement should do better than those that do not. But we shall never know for sure if we do not begin to devise some systematic, well-controlled experiments and find out.

And, of course, there is nothing that says that every part of an entire course has to be taught in exactly the same way that every other part is taught. Indeed, as Professor Haley reported, Professor Paden at Illinois is now combining television and programmed instruction. Surely other combinations or "mixes" of different techniques for different purposes are possible. And, if enough "hard" results of well-designed experiments begin to accumulate, there is no reason why various "proven packages" of different devices cannot be marketed and sold just as textbooks and other materials are now available. If TV or sound film is proven to have certain advantages for presenting particular aspects of economics, for example, why not tape or film just these aspects, and have a series of "shorts" or "specials" that can be piped into closed-circuit systems for flexible use by different instructors at different times? If overhead projector transparencies can build up certain key diagrams in a way that is far superior to the typical instructor's blackboard work, why not use just those transparencies without reproducing every single diagram and table in the textbook? If certain games or simulation exercises prove particularly valuable, why not "package" the instruction cards, participant materials, and follow-up forms for commercial distribution? And so on.

It seems to me that we now are at a particularly propitious time for carefully rethinking our introductory course objectives, devising carefully controlled experiments to measure the attainment of particular objectives, and using the results of these experiments to assemble a "menu" of proven possibilities that offers real hope of pumping life, variety, and excitement into the important area of introductory economics.

AMERICAN ECONOMIC ASSOCIATION

PROCEEDINGS

OF THE

SEVENTY-NINTH

ANNUAL

MEETING

o

•

SAN FRANCISCO, CALIFORNIA
DECEMBER 27-29, 1966

IN MEMORIAM

James Washington Bell

1890-1966

James Washington Bell, Emeritus Professor of Economics, Northwestern University, and former Secretary-Treasurer of the American Economic Association, died, after a short illness, in Evanston, Illinois, December 21, 1966.

A native of Colorado, Professor Bell attended the University of Colorado, where he received a bachelor's degree in 1912, an M.A. degree in 1913, and was an instructor in the Economics Department between 1912 and 1914. In the fall of 1914, he received a teaching fellowship at Harvard University and in 1918 was awarded a Ph.D. degree in economics. During 1918-19 he served with the American Expeditionary Forces in France, with the rank of Second Lieutenant. Appointed an Assistant Professor of Economics at Williams College in 1919, he moved to Northwestern University in 1922, with the rank of Associate Professor and in 1929 was advanced to full Professor. At Northwestern he combined teaching with administration. He was Acting Dean of the Graduate School from 1937 to 1941 and Chairman of the Economics Department from 1942 to 1951. Following retirement from Northwestern in 1955, he was a visiting professor at Ohio State University in 1956 and Southern Illinois University in 1957 and at the time of his death held the position of Lecturer in Economics at the University of the Americas in Mexico City.

Professor Bell had a deep and continuing interest in the areas of money, banking, and monetary policies. While his administrative responsibilities left little time for research and publication, as a director of two Chicago banks, his advice on banking operations was especially welcome. In addition, he was an active member of the Economists' National Committee on Monetary Policy. Of his many interests and activities, however, the one that undoubtedly gave him the greatest pleasure was his work with the American Economic Association. He had already been a member for twenty years when in 1936 he was selected to succeed Frederick S. Deibler as Secretary-Treasurer, a position which he held until the end of 1961.

During this twenty-five year period, Professor Bell carried out his duties faithfully and with distinction. As an ex officio member of various committees, he played an important role in the policy decisions adopted by the Association. He was especially effective in managing the business affairs of the organization. Throughout his term of office, the services of the Association were made available to members at a moderate cost. All who had occasion to associate with Jim Bell were impressed with his personal integrity, his unfailing good nature, and his willingness to help others.

PROCEEDINGS OF THE
AMERICAN ECONOMIC ASSOCIATION

ANNUAL BUSINESS MEETING, DECEMBER 28, 1966
SAN FRANCISCO HILTON HOTEL, SAN FRANCISCO, CALIFORNIA

The Seventy-ninth Annual Business Meeting of the American Economic Association was called to order in the Teakwood Suite of the San Francisco Hilton Hotel, San Francisco, California, at 9:20 p.m. by President Fritz Machlup. The minutes of the Business Meeting of December 30, 1965, were approved and the minutes of the Executive Committee meetings and reports of officers and committees of the Association were ratified. These reports are published in these "Proceedings" and constitute the official actions of the Association when approved at the annual meeting.

It was noted that James Washington Bell, Secretary-Treasurer of the Association from 1936 through 1961 had passed away on December 21, 1966. It was VOTED to include a memorial to Professor Bell in the 1966 "Proceedings."

The report of the Secretary was presented by H. F. Williamson. The schedule of times and places for future annual meetings through 1973 was outlined. Attention was called to the fact that the total number of members and subscribers as of November 30, 1966, was 21,304, a net increase of 1,478 for the year.

In addition, the Secretary reported on the following activities of the Association:

1. *Clearinghouse Project.* There are currently over 2,500 individuals in the roster of economists interested in overseas assignments.

2. *National Register of Scientific and Technical Personnel.* The Association participated in the 1966 N.S.F. Register of Scientific Personnel under a special arrangement with the N.S.F. Professor Arnold Tolles will prepare a more extensive analysis of the economics profession based on data from the 1964 and 1966 returns from the Register.

3. *Foreign Graduate Student Screening Project.* This project, now in its second year of operation, is prepared to handle requests from economics departments for an evaluation of foreign students from countries in the Middle and Far East, Latin America, and Africa.

4. *Group Flights.* Four group flights were arranged for attendance at the San Francisco meeting, and it is planned to arrange similar flights for future meetings.

5. *United States Employment Service.* The United States Employment Service, which has operated for a number of years at our annual meeting, has arranged to make its services available on a year-round basis.

6. *Handbook.* Upon the recommendation of the Executive Committee, it was decided to publish a comprehensive edition of the *Handbook* in 1968.

The report of the Treasurer was summarized by H. F. Williamson. The net

operating income of the Association for the year ending November 30, 1966, was $29,736 and the unappropriated surplus was $267,628. The full details are shown in the Treasurer's report printed below.

H. F. Williamson reported for the Finance Committee on the changes in the Association's investment portfolio. The full report of the Finance Committee is published below.

The report of the Managing Editor, published below, was summarized by J. G. Gurley. He expressed his appreciation for the effective work done by the Editorial Board and consultants in preparing the *American Economic Review*.

The Secretary then presented the report of the Committee on Elections and the certification of the election of new officers for the year 1967 as follows:

In accordance with the bylaws on election procedure, I hereby certify the results of the recent balloting and present the reports of the Nominating Committee and the Committee on Elections.

The Nominating Committee, consisting of Bernard F. Haley, Chairman, Marvin Frankel, Kermit Gordon, Juanita M. Kreps, Stanley Lebergott, and David W. Slater, presented to the Secretary the list of nominees for the respective offices:

<div align="center">

For President-elect
Kenneth E. Boulding

</div>

For Vice-Presidents	*For Executive Committee*
William J. Baumol	Kenneth J. Arrow
Gerhard Colm	Richard A. Easterlin
John T. Dunlop	Franco Modigliani
Robert Triffin	Joseph A. Pechman

The Committee on Elections, consisting of Irma Adelman, Chairman, Haskel Benishay, and Harold F. Williamson, prepared biographical sketches of the candidates and ballots were distributed early in November. The canvass of the ballots was made on December 14, 1966, and the results were filed with the Secretary.

From the report of the Committee on Elections, I have the following information:

Number of envelopes without name for identification	53
Number received too late	36
Number of defective ballots	—
Number of legal ballots	5,304
Number of returns from the mail ballot	5,393

On the basis of the canvass of the votes cast, I certify that the following persons have been duly elected to the respective offices:

President-elect (for a term of one year)
    Kenneth E. Boulding

Vice-Presidents (for a term of one year)
    William J. Baumol
    Robert Triffin

Members of the Executive Committee (for a term of three years)
    Kenneth J. Arrow
    Franco Modigliani

Following the report of the Committee on Elections, President Milton Friedman took over the chair.

The President recognized Harry G. Johnson, who presented the following resolution, which was adopted:

Be it resolved that the Association offer its best thanks to the Committee on Local Arrangements for its success in facilitating the smooth operation of the San Francisco

meeting and in particular congratulate the Committee on the efficiency in arranging for hotel accommodations and conference registration.

The President then recognized Charles P. Kindleberger, who presented the following resolution, which was adopted:

Be it resolved that the Association indicate its indebtedness to President Fritz Machlup for the efficient conduct of his office and for the preparation of a scholarly and thought-provoking Presidential Address.

Be it further resolved that the Association express its appreciation to its President-elect, Milton Friedman, for arranging a program that covered so many subjects of vital interest to our members.

The meeting was adjourned at 9:50 p.m.

HAROLD F. WILLIAMSON, *Secretary*

MINUTES OF THE EXECUTIVE COMMITTEE MEETINGS

1. Minutes of the spring meeting held in New York City, March 11-12, 1966:

The *first meeting* of the *1966 Executive Committee* was called to order at 9:30 a.m. at the New York Hilton Hotel, New York City, March 11, 1966. The following were present: Fritz Machlup, presiding, Moses Abramovitz, J. M. Buchanan, Milton Friedman, H. G. Johnson, C. P. Kindleberger, L. R. Klein, T. C. Koopmans, R. M. Solow, J. J. Spengler, G. J. Stigler, and H. F. Williamson. Absent were: J. G. Gurley and N. H. Jacoby. Present as members of the Nominating Committee were: B. F. Haley, Chairman, Marvin Frankel, Kermit Gordon, Juanita M. Kreps, Stanley Lebergott, and D. W. Slater. Present as guest was: J. P. Miller.

1. *President's Remarks* (Fritz Machlup). President Machlup outlined the items on the agenda to be covered at the meetings.

2. The minutes of the Executive Committee meeting of December 27, 1965, previously sent out to members were approved.

3. *Report of the Secretary* (H. F. Williamson).

Annual Meetings. A preliminary report indicated that the total net registration at the New York City meeting was 5,271.

Following a discussion of arrangements for the 1966 annual meeting in San Francisco, it was decided: (1) to schedule the meeting for December 27-29; (2) to follow a daily time schedule of two morning sessions, 8:30-10:30 and 10:45-12:30 and one afternoon session, 2:30-4:30; (3) to hold the meeting of the Executive Committee on December 26, starting at 3:00 p.m. To insure adequate hotel and meeting room accommodation, the Secretary was authorized to canvass the membership in order to determine the approximate number planning to attend the meeting in San Francisco.

Papers and Proceedings. It was announced that the plan was to print 22,500 copies of the 1965 *Papers and Proceedings* and that the total number of manuscript pages would be approximately 1,120.

Clearinghouse Project. The activities of the Clearinghouse have continued to expand and during the two-year period ending September 30, 1965, the Association has been asked to recommend candidates for approximately 200 positions overseas. Because the Ford Foundation grant to establish the Clearinghouse Project will terminate as of August, 1966, the Association is negotiating with the State Department-AID for financial support to enable the Clearinghouse Project to continue to operate.

N.S.F. Register of Scientific Personnel. Plans have been fully worked out for the Association's participation in the 1966 Register of Scientific Personnel. It was further noted that as soon as an advisory committee is appointed, Arnold Tolles will draw up a proposal to be submitted to N.S.F. for further analysis of the 1964 and 1966 data on economists.

Foreign Student Screening Project. As of March 9, 1966, the Foreign Student Screening Project had received requests from thirteen departments to arrange interviews or evaluate the credentials of forty-five students located in twelve countries outside the United States. In the discussion of this report, it was agreed that despite the small number of requests, the results of one year's operation of the Project were not an adequate test of the potential value of the service involved. It was further agreed that the Secretary should ask representative departments to suggest how the services of the Project can be improved.

Association History Project. G. J. Stigler, Joseph Dorfman, and H. F. Williamson have been appointed to the committee to advise A. W. Coats in the preparation of the history of the Association.

4. *Report of the Treasurer* (H. F. Williamson). The Treasurer presented a budget estimate for fiscal 1966. In connection with the budget, he suggested that except for the years the *Handbook* is published, consideration should be given to the possibility of issuing an annual directory with the names and addresses of the members of the Association. Following a discussion, it was VOTED to appoint Fritz Machlup, Milton Friedman, and H. F. Williamson to an *ad hoc* committee to investigate the cost and feasibility of issuing an annual directory.

5. *Accounting and Investment Policies of the Association.* Milton Friedman indicated that the following considerations had prompted his suggestions, made at the December,

1965, meeting of the Executive Committee, for changes in the accounting and financial policies of the Association: (1) The Association's unappropriated surplus is fully adequate to meet any foreseeable financial emergencies. (2) It is therefore inappropriate for the Association to operate in a way that will continue to add to the surplus account. (3) To get a more accurate measure of the financial operations of the Association the annual income accounts should, in addition to dividends declared, include the undistributed earnings on stocks owned by the Association. (4) The liabilities of the Association are such that, except for current operating expenses and special grants, there is no reason for the Association to invest its surplus in corporate bonds, government securities, or savings accounts.

Following a discussion, it was VOTED that the policy of the Association shall be to invest all funds not needed for current operating expenses or as reserves for special grants in equity holdings.

It was also VOTED to have the President appoint a committee to consider the adoption of accounting procedures that would provide an accurate measure of the income of the Association.

6. *Report of the Managing Editor of the A.E.R.* (J. G. Gurley). In the absence of J. G. Gurley, B. F. Haley presented the Editor's nominees for the *A.E.R.* Editorial Board. It was VOTED to accept the nominations.

7. *Report of the Editor of the Journal of Economic Abstracts* (Arthur Smithies). In the absence of the Editor, the President reported that final arrangements had been worked out to transfer the list of periodical articles from the *A.E.R.* to the *Journal.*

8. *Reports of Standing and Special Committees.*

Committee on Research and Publications (I. B. Kravis). In the absence of the Chairman, H. G. Johnson reported that the "Readings Series" volume on international trade was well along and that he had made arrangements with authors to prepare surveys of national economic problems and policy issues.

Committee on Honors and Awards (B. F. Haley). The Chairman reported that the Committee was not recommending any candidate for the John Bates Clark medal. He also reported on the candidates recommended by the Committee for the Distinguished Fellow award.

Nominating Committee (B. F. 'Haley). The Executive Committee and the Nominating Committee met as an Electoral College to consider nominees for the office of President-elect for 1967. After discussion, the nominee was selected and his acceptance obtained. Nominations for the other offices were discussed. The Chairman also submitted the Committee's recommendations for the Distinguished Fellow award. The recipient was selected as a result of the ballot of the members of the Electoral College.

9. *Unfinished and Miscellaneous Business.*

Index of Economic Journals. J. P. Miller reported that the publication of Volume VII was scheduled for the fall of 1966 and that the preparation of volume VIII was well under way.

Index of Festschrifts. J. P. Miller outlined the problems involved in the preparation of an index of festschrifts. Following a discussion, it was agreed that he should explore the feasibility of the project further and consult with the Committee on Research and Publications on the possibility of preparing such an index.

Group Insurance. The Secretary outlined the proposal to make group, accident, and life insurance available to the members of the Association and their families. Following a discussion, it was decided that it was inconsistent with the purposes of the Association to provide this type of service to the membership.

Association Counsel. At the request of the President, several names were suggested as candidates for appointment as Association counsel, with the understanding the appointment should be made by the President.

Advisory Committee on Government Programs in the Behavioral Sciences. It was VOTED to authorize the President to suggest representatives to serve on the Advisory Committee on Government Programs in the Behavioral Sciences.

Program for the 1966 Meeting. President-elect Milton Friedman discussed his plans for the 1966 annual meeting.

The meeting was adjourned at 12:00 m. on March 12, 1966.

2. Minutes of the meeting held in San Francisco, California, December 26, 1966:

The *second meeting* of the *1966 Executive Committee* was called to order at 3:00 p.m. at the San Francisco Hilton Hotel, Fritz Machlup, presiding. Others present were: Moses

Abramovitz, J. M. Buchanan, Milton Friedman, J. G. Gurley, N. H. Jacoby, H. G. Johnson, C. P. Kindleberger, L. R. Klein, J. J. Spengler, G. J. Stigler, and H. F. Williamson. Absent were: T. C. Koopmans and R. M. Solow. Present as guests were: G. L. Bach, W. J. Baumol, F. M. Boddy, K. E. Boulding, Sidney Davidson, K. A. Fox, W. F. Owen, Simon Rottenberg, and Richard Ruggles.

1. *President's Remarks* (Fritz Machlup). The President reviewed the order of the items on the agenda.

2. *Minutes.* The minutes of the March 11-12, 1966, meeting were approved.

3. *Report of the Secretary* (H. F. Williamson). The Secretary presented the schedule of future meetings, membership growth and composition, publication costs of the 1966 *Papers and Proceedings,* requests to reprint, the use of the mailing list, and new activities of the Association—all more fully treated in the Secretary's report as presented below.

It was VOTED to approve and accept the Secretary's report.

It was VOTED to hold the spring meeting of the Executive Committee in New York City, March 10-11, 1967.

4. *Report of the Treasurer* (H. F. Williamson). The Report of the Treasurer, published in full below, was summarized. Special attention was called to the fact that the net operating income for the fiscal year ending November 30, 1966, was $29,736 and the unappropriated surplus was $267,628.

5. *Report of the Auditor* (H. F. Williamson). The Secretary distributed copies of the auditor's report as prepared by Arthur Andersen & Co. It was VOTED to accept the report of the auditor as printed below. The Secretary was instructed to express the appreciation of the Executive Committee to Arthur Andersen & Co. for their services.

6. *Report of the Finance Committee* (H. F. Williamson). The Secretary outlined the main features of the Finance Committee report, printed in full below. It was VOTED to accept the report and to reappoint C. Wells Farnham and Corliss D. Anderson as members of the Finance Committee for 1967. The Secretary was instructed to express the thanks of the Executive Committee to the members of the Finance Committee for their services during 1966.

7. *American Economic Review* (J. G. Gurley). The Editor reviewed the report on the operations of the Editorial Office of the *American Economic Review* for 1966, published in full below.

It was VOTED to: (a) approve the excess of expenditures by the Editorial Office in 1966 over the budget amount authorized by the Executive Committee; (b) accept the proposed budget for 1967 of $25,725 for the operation of the Editorial Office, plus $1,800 to contributors; and (c) authorize the Editor and Treasurer to pay the costs of printing and distributing a maximum total of 1,420 net pages of the *Review* during 1967. The Editor was requested to make a cost analysis per net page of the *Review* for consideration at the December meeting of the Executive Committee.

8. *Journal of Economic Abstracts* (Arthur Smithies). In the absence of the Editor, Milton Friedman outlined tentative plans for changing the content of the *Journal.* Following a discussion of these plans, it was VOTED to: (a) authorize a budget of $1,000 for the operation of the Editorial Office of the *Journal* for the period January 1-March 10, 1967; (b) ask the Editor of the *Journal* to prepare a report and a budget for 1967 for approval by the Executive Committee at the spring meeting.

9. *Reports of Standing and Special Committees.*

Committee on Research and Publications (I. B. Kravis). Report printed below.

Index of Economic Essays Published in Festschriften and Symposia (Richard Ruggles). The Chairman outlined the plans for the preparation of the *Index of Economic Essays Published in Festschriften and Symposia.* Following a discussion, it was VOTED to make an advance of $2,000 to Yale University to finance the costs of preparing the *Index.*

Committee on Economic Education (G. L. Bach). In summarizing his report on the activities of the Committee on Economic Education, published below, the Chairman noted the extent to which Committee members have become increasingly involved in representing the profession, particularly in contacts with federal, state, and local governmental educational agencies. He suggested that if the profession is to be adequately represented in the activities of these agencies, it may be necessary to work out an arrangement that will not be an undue burden on Committee members.

Committee on Financial and Investment Policies (Milton Friedman). Sidney Davidson outlined the Committee's report, published in full below. He called special attention to the recommendation of the Committee for the adoption of an accounting method by the Association that would "make the Association's income from equities in each year equal to dividends received in that year plus our share of earnings plowed back that year plus one-third of market appreciation in that year (adjusted for retained earnings and price level

change) plus one-third of adjusted appreciation in the preceding year, plus one-third of adjusted appreciation in the second preceding year." He also noted that in the judgment of the Committee, the adoption of the proposed formula for reporting income by the Association might well set a precedent for other associations and institutions to follow.

Following the discussion, it was VOTED to: (a) change the fiscal year of the Association from December 1-November 30 to the calendar year, starting with January 1, 1967, and (b) adopt the accounting method proposed by the Committee in reporting all equity income of the Association.

Joint Committee of the American Economic Association and Association of American Law Schools (G. J. Stigler). G. J. Stigler summarized the report of the Joint Committee, printed below. Following discussion, it was VOTED to approve the plan recommended in the report for the Association to cooperate with the A.A.L.S. in the preparation of monographs on the relations between law and economics.

10. *Reports of A.E.A. Representatives or Advisory Committees.*

A.C.L.S. (R. L. Andreano). No report.

S.S.R.C. (K. A. Fox). Karl Fox reported informally on the activities of the S.S.R.C.

N.B.E.R. (W. L. Thorp). Report printed below.

A.A.A.S. (Carl Kaysen). No report.

I.E.A. (Fritz Machlup). No report.

I.I.E. (B. H. Higgins). On behalf of the Board of the Institute, Simon Rottenberg and Wyn Owen requested the Association's endorsement of the Institute's application to the Ford Foundation for a grant to supplement its activities. Following a discussion, it was VOTED to endorse the Institute's application to the Ford Foundation.

Census Advisory Committee (Solomon Fabricant). Report printed below.

UNESCO (K. E. Boulding). No report.

Joint Council on Economic Education (G. L. Bach). See report of Committee on Economic Education.

National Research Council, National Academy of Science (Lionel McKenzie). See report printed below.

J.C.E.E. Special Advisory Board (G. L. Bach). See report of the Committee on Economic Education.

11. *Unfinished and Miscellaneous Business.*

Francis Boddy outlined a proposal by the Council of Graduate Schools for a pilot study of the feasibility of allocating financial aid to graduate students in economics similar to the matching procedures followed by American medical schools in allocating internships. Following a discussion, it was agreed that when the details of the proposal had been worked out, the Association would consider the possibility of endorsing the plan on behalf of the profession.

J. G. Gurley reported on a letter from the publisher of *Surveys of Economic Theory,* asking how royalties on the sale of the volumes should be distributed. It was VOTED to request the publisher to turn over the appropriate share of royalties to the Association.

It was VOTED to appropriate annually an amount of $1,000 both to the president and the president-elect to cover the expenses incurred in connection with their duties as officers of the Association.

After a consideration of the advantages of having the results of the annual election of officers available before the latter part of December, it was VOTED to mail out the ballots in September.

In response to the question raised by the Secretary regarding future editions of the handbook, it was VOTED to consider the possibility of issuing a "who's who" type of handbook every four years, starting in 1968, and a membership-list type of handbook every four years, starting in 1970.

The meeting was adjourned at 10:40 p.m.

REPORT OF THE SECRETARY FOR THE YEAR 1966

Annual Meetings. The final report for the 1965 annual meeting in New York City indicated a total registration of 5,271. Of this number, 3,593 were members of the American Economic Association. The net income from the meeting amounted to $26,104.47. The Association's share of the total, based on the number of members who registered, was $17,360.94.

The schedule for future annual meetings is: 1967, Washington, D.C., Shearton-Park Hotel; 1968, Chicago, Pick-Congress Hotel; 1969, New York City, New York Hilton Hotel; 1970, Detroit, Sheraton-Cadillac Hotel; 1971, New Orleans (provided adequate hotel accommodations are available); 1972, Chicago, Conrad Hilton Hotel; 1973, New York City, New York Hilton Hotel.

Dr. Gerald Pollack, of the U.S. Department of Commerce, will serve as Chairman of the Committee on Local Arrangements for the 1967 annual meeting in Washington, D.C.

Membership. Exhibit I below shows that as of November 30, 1966, the total number of members and subscribers was 21,304, a net increase of 1,478 for the year.

Advertising and Announcements. The number of advertising pages in the *American Economic Review* in 1966 was 183½ paid and 52½ exchange pages compared to 167½ paid and 51 exchange pages in 1965. The "Vacancies and Applications" announcements totaled 47 pages in 1966 compared to 35 pages in 1965.

Papers and Proceedings. A comparison of the size and cost of the *Papers and Proceedings* and the *Handbooks* for the period 1957-66 is shown in Exhibit II.

Permission to Reprint and Translate. Official permission to quote from, reprint, or translate articles from the *Review* and the *Papers and Proceedings* totaled 215 in 1966 compared to 161 in 1965. Of the 215 permissions granted in 1966, 95 were in response to requests to reproduce or reprint articles for classroom or conference use.

Use of Mailing List. Requests for the use of the Association's mailing list continue to come largely from publishers of books and periodicals and from government agencies, corporations, and labor organizations wishing to send out reports or reprints of speeches. It is the policy of the Association to grant requests for the use of the mailing list only when the material to be distributed appears to be of interest to a substantial proportion of our members. Income from the sale of the mailing list in 1966 amounted to approximately $7,500.

Clearinghouse Project. There are currently nearly 2,500 economists interested in overseas assignments registered with the American Economic Association Clearinghouse Project. Between October, 1963, when the Project began operating, and December, 1966, some 100 organizations have called on the

Project to suggest candidates for approximately 200 overseas positions. The most frequent users of the service are: the United States State Department, particularly AID and the International Recruitment Service, the United Nations, Overseas Educational Service, and the Ford Foundation. It appears quite likely that a contract will be negotiated between the Association and AID that will insure the financial support necessary for a continued operation of the Clearinghouse Project.

Foreign Students Screening Project. Under the Foreign Students Screening Project the Association will continue to provide evaluations of candidates from the major countries in the Middle East, Far East, Latin America, and Africa who apply for admission as graduate students to American and Canadian departments of economics and agricultural economics.

National Register of Scientific Personnel. Under a contract with the National Science Foundation, the Association participated for the second time during 1966 in the survey of scientific personnel in the United States. The data on the profession available from the survey will be used by Arnold Tolles to update and expand the analysis of the *Structure of Economists' Employment and Salaries, 1964,* published as a supplement to the December, 1965, *American Economic Review.*

Employment Service. Following a long-established practice, the U.S. Employment Service arranged to operate an employment register at the annual meeting in San Francisco. The growing use of the service is suggested by the following data on its operation at the annual meeting in New York in 1965: number of job openings listed, 1,836; number of applicants, 1,488; number of messages handled, 11,500.

In addition, arrangements have been made with the U.S. Employment Service to provide a year-round service for employers and job applicants through the National Registry for Economists, 208 South La Salle Street, Chicago, Illinois, 60604.

Group Flights. There was sufficient interest among members of the Allied Social Science Associations to arrange group flights to the annual meeting in San Francisco from Boston, New York City, Washington, D.C., and Chicago. In cities such as Detroit, Minneapolis, Pittsburgh, and St. Louis, where less than the required minimum number of twenty-five signed up for group flights. reservations were made available to members at the normal fare. Present plans are to arrange for group flights to the annual meeting in Washington, D.C., in 1967

Standing Committees

CENSUS ADVISORY COMMITTEE
  Solomon Fabricant, *Chairman* (1968)
  Bert G. Hickman (1966)
  Guy H. Orcutt (1966)
  Sherman J. Maisel (1967)
  Ralph W. Pfouts (1967)
  Gideon Rosenbluth (1967)
  Daniel Creamer (1968)
  H. Gregg Lewis (1968)
  John Lintner (1968)
  Arthur M. Okun (1968)
  Morris A. Adelman (1969)
  Robert R. Nathan (1969)
  Fred H. Klopstock (1969)
  Ernest Williams, Jr. (1969)
  Douglass C. North (1969)

COMMITTEE ON ADMINISTRATIVE
    STAFF SALARIES
  Joseph J. Spengler
  Fritz Machlup
  Milton Friedman

COMMITTEE ON ECONOMIC
    EDUCATION
  G. L. Bach, *Chairman* (1967)
  Marshall R. Colberg (1967)
  Rendigs Fels (1967)
  Emanuel T. Weiler (1967)
  Kenneth E. Boulding (1969)
  Robert A. Gordon (1969)
  Ben W. Lewis (1969)
  Harold F. Williamson, Ex Officio

COMMITTEE ON HONORS AND AWARDS
  Bernard F. Haley, *Chairman*
    (1966)
  Earl J. Hamilton (1966)
  Martin Bronfenbrenner (1966)
  James Tobin (1970)
  Anthony D. Scott (1970)
  Richard A. Musgrave (1970)

COMMITTEE ON RESEARCH AND
    PUBLICATIONS
  Irving B. Kravis, *Chairman* (1966)
  Karl A. Fox (1967)
  Franco Modigliani (1968)
  Willard L. Thorp (1970)
  Bert F. Hoselitz (1966)
  Harry Johnson (1966)
  Harold F. Williamson, Ex Officio

INSTITUTE OF INTERNATIONAL EDU-
    CATION ADVISORY AND POLICY
    BOARD
  Benjamin Higgins, *Chairman*
    (1966)
  Simon Rottenberg (1967)
  Anthony Tang (1966)
  Raymond Mikesell (1967)
  Henry Rosovsky (1967)
  Neil W. Chamberlain (1968)
  H. Gregg Lewis (1968)

JOINT COUNCIL ON ECONOMIC EDU-
    CATION, SPECIAL ADVISORY
    COMMITTEE
  G. L. Bach (1967)
  Rendigs Fels (1967)
  Ben W. Lewis (1969)

TECHNICAL SUBCOMMITTEE ON BUSI-
    NESS CYCLE DEVELOPMENTS
  Bert G. Hickman, *Chairman*
    (1966)
  Donald J. Daly (1967)
  Gottfried Haberler (1966)
  Lawrence R. Klein (1967)
  Geoffrey H. Moore (1968)
  Frank E. Morris (1969)
  Arthur M. Okun (1967)
  Beryl W. Sprinkel (1969)
  Lorman C. Trueblood (1966)
  Harry Stark (1969)

Committees Appointed During the Year

ADVISORY COMMITTEE ON THE HIS-
    TORY OF THE ASSOCIATION
  George J. Stigler, *Chairman*
  Harold F. Williamson
  Joseph Dorfman

COMMITTEE ON THE INDEX OF
    ECONOMIC JOURNALS
  Richard Ruggles, *Chairman*

ADVISORY COMMITTEE ON THE STUDY
    OF THE STRUCTURE OF THE
    PROFESSION
  Karl A. Fox, *Chairman*
  William J. Baumol
  Francis M. Boddy
  Gary S. Becker

COMMITTEE ON THE AMERICAN ECO-
NOMIC REVIEW AND THE JOUR-
NAL OF ECONOMIC ABSTRACTS

Milton Friedman, *Chairman*
Bernard F. Haley
Neil H. Jacoby
Robert H. Strotz
Fritz Machlup
Kenneth E. Boulding
Harold F. Williamson, Ex Officio

COMMITTEE ON CLASSIFICATION

Richard H. Leftwich, *Chairman*
Richard Ruggles
John G. Gurley
Arthur Smithies
Harold F. Williamson

COMMITTEE ON ELECTIONS

Irma Adelman, *Chairman*
Haskel Benishay
Harold F. Williamson, Ex Officio

COMMITTEE ON FINANCIAL AND
INVESTMENT POLICIES

Milton Friedman, *Chairman*
Robert Eisner
Sidney Davidson
C. Wells Farnham
Harold F. Williamson

COMMITTEE ON UNESCO

Kenneth E. Boulding, *Chairman*
James Blackman
Jerry Miner

FINANCE COMMITTEE

C. Wells Farnham, *Chairman*
Corliss D. Anderson
James Washington Bell
Harold F. Williamson

JOINT COMMITTEE WITH THE ASSOCIA-
TION OF AMERICAN LAW SCHOOLS

Armen A. Alchian
Ronald H. Coase
Mark S. Massel
Gerald M. Meier
George J. Stigler

NOMINATING COMMITTEE

Bernard F. Haley, *Chairman*
Marvin Frankel
Kermit Gordon
Juanita M. Kreps
Stanley Lebergott
David W. Slater

Council and Other Representatives

A.A.A.S.

Carl Kaysen (1968)

A.C.L.S.

Ralph L. Andreano (1966)

NATIONAL RESEARCH COUNCIL OF
THE NATIONAL ACADEMY OF
SCIENCES

Lionel McKenzie (1967)
James N. Morgan (1968)
James Tobin (1969)

N.B.E.R.

Willard L. Thorp (1970)

NATIONAL COUNCIL OF TESTING OF
ENGLISH AS A FOREIGN
LANGUAGE

Arthur E. Burns

I.E.A.

John T. Dunlop (1968)
Fritz Machlup (1970)

N.S.F. INTERDEPARTMENTAL PANEL
ON SCIENTIFIC INFORMATION
ACTIVITIES

Richard Ruggles

S.S.R.C.

G. H. Hildebrand (1966)
Karl A. Fox (1967)
Franco Modigliani (1968)

U.S. NATIONAL COMMISSION FOR
    UNESCO
Kenneth E. Boulding

Representatives of the Association on Various Occasions

OFFICE OF SCIENCE AND TECHNOLOGY (OFFICE OF THE PRESIDENT, WASH-
    INGTON, D.C.), PANEL ON PROTECTION OF PRIVACY IN FEDERALLY SPON-
    SORED BEHAVIORAL RESEARCH
    Gerhard Bry
AMERICAN ACADEMY OF POLITICAL AND SOCIAL SCIENCE SEVENTIETH ANNUAL
    MEETING
    Karl R. Bopp and Irving B. Kravis
CALIFORNIA INSTITUTE OF TECHNOLOGY 75TH ANNIVERSARY CONVOCATION
    Harold M. Somers
BOSTON COLLEGE, DEDICATION OF HIGGINS HALL AND UNIVERSITY CONVOCA-
    TION
    Everett E. Hagen
CONFERENCE OF POPULATION AND HOUSING CENSUS USERS
    Sam H. Schurr
ALLIED SOCIAL SCIENCE COMMITTEE ON FEDERAL STATISTICS (BUREAU OF
    THE BUDGET)
    Sam H. Schurr

INAUGURATIONS
Newton Edd Miller, Reno Campus, University of Nevada
    Ben W. Wofford
John E. Champion, Florida State University
    John B. McFerrin
William Robert Parks, Iowa State University
    Karl A. Fox
Joseph P. McMurray, Queens College
    Arnold W. Sametz
John S. Toll, State University of New York at Stony Brook
    Albert G. Hart
Glenn Leggett, Grinnell College
    Robert Voertman
Wilbert E. Locklin, Springfield College
    Marshall C. Howard
David Alexander, Southwestern at Memphis
    McDonald K. Horne, Jr.
Harry Melvin Philpott, Auburn University
    Ralph M. Havens
James Everett Purdue, State University College at Oswego
    Richard Hyse
Stanley Heywood, Eastern Montana College
    Clifford H. Clark

Carl Gustaf Fjellman, Upsala College
   Henry F. Schoenbeck
Norton Nelson, Provost for University Heights, New York University
   N. T. Wang
Howard Wesley Johnson, Massachusetts Institute of Technology
   Don D. Humphrey
Albert Edward Holland, Hobart and William Smith Colleges
   Chandler Morse
Robert A. Christie, Millersville State College
   Will Lyons
Ray Lorenzo Heffner, Brown University
   Jerome B. Stein
Francis Rawle Haig, Wheeling College
   John W. May
Malcolm Carron, University of Detroit
   Thomas Finn
Terrence J. Murphy, College of Saint Thomas
   Norman J. Simler
Paul F. Sharp, Drake University
   Karl A. Fox
Walter Raleigh Coppedge, College of Charleston
   Robert Spencer Adden
Grover E. Murray, Texas Technological College
   Vernon Thomas Clover
William Henry Masterson, University of Chattanooga
   John R. Moore
Albert LeRoy Pugsley, Youngstown University
   Emile Grunberg
Fred Blake Bentley, Mars Hill College
   Roy Wood
Elizabeth J. McCormack, Manhattanville College of the Sacred Heart
   Walter A. Chudson
Albert W. Brown, State University College at Brockport
   Harold W. Miller

                        Respectfully submitted
                        HAROLD F. WILLIAMSON, *Secretary*

EXHIBIT I

MEMBERS AND SUBSCRIBERS

	Totals 11/30/65	Gain or *Loss*	Totals 11/30/66
Class of membership:			
Annual................	11,902	918	12,820
Junior.................	1,677	142	1,819
Family................	184	12	196
Complimentary..........	100	23	123*
Life..................	247	7	254
Honorary..............	17	—	17
Total members........	14,127	1,102	15,229
Subscribers..............	5,699	376	6,075
Totals................	19,826	1,478	21,304

* Includes 49 who do not receive the publications.

EXHIBIT II

PUBLICATION COSTS

Year*	PAPERS AND PROCEEDINGS			HANDBOOKS		
	Number of pages	Number of copies	Cost	Number of pages	Number of copies	Cost
1957	754	12,400	$16,253	548	10,100	$15,815
1958	677	12,700	15,471	32	9,300	1,434
1959	689	14,000	16,780			
1960	745	14,800	18,914			
1961	675	15,800	19,759			
1962	615	16,000	18,277			
1963	753	17,700	23,440			
1964	698	18,500	23,362	472	16,000	48,626
1965	652	20,000	23,264			
1966	670	22,500	28,405	184	23,000	13,261

* This is the year of publication and pertains to the meeting of the preceding year.

REPORT OF THE TREASURER
FOR THE YEAR ENDING NOVEMBER 30, 1966

The comparative financial results of the operation of the Association for 1965 and 1966 are shown in detail in the Auditor's Report, Exhibits 3 and 4, printed below. While receipts from sales of back numbers of the *American Economic Review* and the *Handbook* and the mailing list were lower in 1966 than in 1965, the returns from other sources, particularly membership dues, subscriptions, and the annual meeting expanded sufficiently to raise total operating income in 1966 to $301,053 compared to $244,586 in 1965. General administrative expenses were lower in 1966 than in 1965, but there was a sharp increase in the costs of publication, in part the result of the publication of the 1966 *Handbook* but more importantly attributable to the expense of printing and distributing the *Journal of Economic Abstracts* to members and subscribers. Total operating expenses for 1966 were $271,317 compared to $213,092 for 1965. The overall result was a net operating income of $29,736 for 1966 compared to $31,494 for 1965.

A detailed comparison of the Association's financial status for 1965 and 1966 is also shown in the Auditor's Report, Exhibit 1. Largely as the result of adding net operating income to the surplus account, the General Fund Balance or Unappropriated Surplus of the Association at the end of 1966 was $267,628 compared to $237,892 at the end of 1965.

Tables 1 and 2 give summaries of the investment portfolios and return on investment for the years 1945-66. A list of the Association's security holdings for 1965 and 1966 is shown in Table 1 of the Report of the Finance Committee, printed below.

Respectfully submitted,
HAROLD F. WILLIAMSON, *Treasurer*

TABLE 1

INVESTMENT PORTFOLIO

| YEAR | AT PAR | COST | | | MARKET |
	Bonds	Bonds	Stocks	Total	Stocks and Bonds
1945	$ 40,000	$ 36,705	$ 44,955	$ 81,661	$103,574
1948	35,000	33,108	48,624	81,732	84,841
1950	35,000	33,108	51,978	85,087	104,177
1951	43,000	43,340	49,764	93,104	117,316
1952	42,000	42,312	58,934	101,246	130,836
1953	68,000	68,308	46,458	114,766	134,562
1954	61,000	61,518	38,082	99,600	132,280
1955	75,000	75,370	59,394	134,764	166,772
1956	75,000	75,370	60,237	135,607	168,337
1957	75,000	75,370	55,084	130,454	151,638
1958	75,000	75,370	67,741	143,111	175,609
{1959	75,000	75,386	67,652	143,038	191,506
{1959*	175,000	175,616	67,652	243,268	291,506
1960*	160,000	160,508	94,910	255,418	299,768
1961*	170,000	169,794	109,071	278,865	356,131
1962*	125,000	125,367	116,699	242,066	293,039
1963*	90,000	90,367	119,524	209,891	284,160
1964*	150,000	149,198	131,249	280,447	371,556
1965*	125,000	124,048	157,836	281,884	409,653
1966*	143,000	142,138	241,075	383,213	491,324

* Includes bonds held in temporary operating fund.

TABLE 2

RETURN ON INVESTMENTS

Year	Bonds	Stocks	Total	Rate of Return on Cost
1945	$1,479	$2,488	$3,968	4.71%
1948	1,194	2,944	4,139	5.06
1950	1,117	3,860	4,977	5.85
1951	1,026	4,607	5,633	6.05
1952	1,117	3,681	4,799	4.75
1953	1,435	3,587	5,022	4.36
1954	1,621	2,961	4,582	4.58
1955	1,750	3,002	4,752	3.53
1956	1,770	3,336	5,106	3.76
1957	1,770	3,397	5,167	3.90
1958	1,770	3,182	4,952	3.46
{1959*	2,518	3,231	5,749	3.90
{1959†	3,894	3,231	7,125	2.90
1960†	6,693	3,772	10,465	4.09
1961	5,460	4,143	9,603	3.44
1962	4,838	4,489	9,327	3.85
1963‡	3,320	5,041	8,361	3.98
1964‡	3,341	6,364	9,705	3.46
1965‡	5,286	6,330	11,616	4.12
1966‡	6,356	7,957	14,312	3.73

* Does not include income from bonds held in temporary operating fund.
† Includes income from bonds held in temporary operating fund.
‡ Does not include interest on savings account.

TABLE 1 (*Continued*)

	Par or Shares	Market Value	Cost
Electrical Products			
Motorola................................	150	15,000	6,745
Westinghouse Electric......................	200	10,600	11,615
Zenith Radio............................	300	15,600	1,975
		41,200*	
Office Equipment			
International Business Machines............	59	22,184	9,870
Xerox..................................	50	10,250	5,305
		32,434	
Miscellaneous			
Disney Productions.......................	219 39/100	14,235	8,298
Donnelley, R. R..........................	300	11,100	9,795
Jostens.................................	1000	11,000	15,220
		36,335*	
Foreign			
Bayer, A. G.............................	150	5,100	6,820
Siemens, A. G...........................	110	4,730	5,520
Swedish Ball Bearing.....................	200	10,000	14,525
		19,830*	
TOTAL COMMON STOCKS..................		336,574	
TOTAL SECURITIES......................		458,324	
TEMPORARY OPERATING FUND			
U.S. Treas. Bills 12/1/66..................	$33,000		

TABLE 2

SUMMARY OF SECURITIES PURCHASED AND SOLD
YEAR ENDED NOVEMBER 30, 1966

	Shares	Cost	Proceeds	Gain or *Loss*
Purchases—Stocks				
IBM Corp...................	2	$ 604.42		
Royal Dutch Petroleum........	200	7,424.76		
Walt Disney Productions				
(fractional share)............	1	49.58		
Swedish Ball Bearings..........	116	5,502.46		
Jostens, Inc. (33⅓ par).........	1,000	15,220.00		
Westinghouse Electric Corp.....	200	11,614.52		
First Bank Stock Corp.........	400	12,689.01		
Eastern Airlines, Inc..........	100	8,947.90		
United Airlines, Inc............	100	6,107.56		
R. R. Donnelley & Sons........	300	9,792.95		
Marathon Oil Co..............	100	5,632.09		
Sales—Stocks				
Siemens and Halske (rights)....			$346.50	
Stock Dividends and Splits				
IBM (3 for 2 split)...........	19			
Rex Chainbelt, Inc. (2 for 1 split)	150			
Walt Disney Productions				
(3% dividend)..............	6			
Zenith Radio Corp. (2 for 1 split)	150			
		$83,585.25	$346.50	

REPORT OF THE AUDITOR

To the Executive Committee of
American Economic Association:

We have examined the statement of assets and liabilities of American Economic Association (a District of Columbia corporation, organized not for profit) as of November 30, 1966, and the related statements of changes in restricted fund balances and income and expenses for the year then ended. Our examination was made in accordance with generally accepted auditing standards, and accordingly included such tests of the accounting records and such other auditing procedures as we considered necessary in the circumstances.

In our opinion, the accompanying statements present fairly the assets and liabilities of American Economic Association as of November 30, 1966, and the changes in restricted fund balances and income and expenses for the year then ended, in conformity with generally accepted accounting principles applied on a basis consistent with that of the preceding year.

ARTHUR ANDERSEN & CO.

Chicago, Illinois,
December 14, 1966.

EXHIBIT 1

AMERICAN ECONOMIC ASSOCIATION
STATEMENT OF ASSETS AND LIABILITIES
NOVEMBER 30, 1966 AND 1965

Assets	1966	1965
CASH:		
Checking accounts	$ 43,493	$ 46,469
Savings account	—	80,203
	$ 43,493	$126,672
MARKETABLE SECURITIES, at cost:		
United States Government obligations (quoted market price of $139,950 in 1966 and $121,854 in 1965)	$142,138	$124,048
Corporate stocks (quoted market price of $351,374 in 1966 and $287,799 in 1965)	241,075	157,836
	$383,213	$281,884
ACCOUNTS RECEIVABLE, less reserve of $1,400 in 1966 and $800 in 1965	$ 31,963	$ 22,134
ADVANCES, relating to United States Government grant for foreign student screening and selection	$ 9,300	$ 11,473
PREPAID EXPENSES	$ 555	$ 1,615
OFFICE FURNITURE AND EQUIPMENT, at cost, less reserve for depreciation of $2,221 in 1966 and $1,720 in 1965	$ 4,887	$ 3,953
	$473,411	$447,731

Liabilities	1966	1965
ACCOUNTS PAYABLE	$ 30,528	$ 22,327
RESERVE FOR HANDBOOK	$ 24,000	$ 16,000
DEFERRED INCOME:		
Life membership dues	$ 28,200	$ 26,800
Prepaid membership dues	11,378	6,720
Prepaid subscriptions	24,605	20,054
	$ 64,183	$ 53,574
RESTRICTED FUND BALANCES (Exhibit 2)	$ 87,072	$117,938
GENERAL FUND BALANCE:		
Balance, beginning of year	$237,892	$206,398
Excess of income over expenses for the year (Exhibit 3)	29,736	31,494
Balance, end of year	$267,628	$237,892
	$473,411	$447,731

EXHIBIT 2

AMERICAN ECONOMIC ASSOCIATION
STATEMENT OF CHANGES IN RESTRICTED FUND BALANCES
FOR THE YEAR ENDED NOVEMBER 30, 1966

Fund	Balance Nov. 30, 1965	Receipts	Disburse-ments	Balance Nov. 30, 1966
The Ford Foundation grants for—				
Translation of foreign economic publications and survey of foreign economic research............................	$ 40,989	$ —	$ 5,824	$35,165
Publication of the *Journal of Economic Abstracts*.........................	1,226	—	1,226	—
Clearinghouse for economists interested in overseas assignments............	11,567	902	10,420	2,049
Foreign student screening and selection	48,340	1,319	14,552	35,107
United States government grant for foreign student screening and selection........	13,008	—	3,073	9,935
The Carnegie Corporation of New York grant for travel expenses of delegates to international meetings...............	382	—	382	—
The Asia Foundation for Asian economist membership dues to the American Economic Association and related travel expenses.............................	1,648	2,500	1,634	2,514
The Agricultural Development Council, Inc.................................	667	—	347	320
The National Science Foundation grant for visiting scientist program.............	—	2,050	223	1,827
Sundry grants.........................	111	100	56	155
	$117,938	$6,871	$37,737	$87,072

EXHIBIT 3

AMERICAN ECONOMIC ASSOCIATION
STATEMENT OF INCOME AND EXPENSES
FOR THE YEARS ENDED NOVEMBER 30, 1966 AND 1965

	1966	*1965*
INCOME:		
Membership dues...	$135,302	$102,678
Subscriptions...	62,239	42,726
Advertising...	52,982	47,273
Sale of copies, republications and handbooks....................	8,098	9,783
Sale of mailing list..	7,483	12,164
Annual meeting...	18,088	11,503
Interest and dividends......................................	15,391	14,771
Gain on sale of securities...................................	—	2,192
Sundry..	1,470	1,496
Total income...	$301,053	$244,586
EXPENSES:		
Publication (Exhibit 4)......................................	$200,073	$130,015
Administrative (Exhibit 4)..................................	68,744	78,077
Grant to Yale University for Cumulative *Index of Economic Journals*	2,500	5,000
	$271,317	$213,092
Excess of income over expenses.............................	$ 29,736	$ 31,494

The accompanying statement of publication and administrative expenses is an integral part of the above statement.

EXHIBIT 4

AMERICAN ECONOMIC ASSOCIATION
STATEMENT OF PUBLICATION AND ADMINISTRATIVE EXPENSES
FOR THE YEARS ENDED NOVEMBER 30, 1966 AND 1965

	1966	1965
PUBLICATION EXPENSES:		
Printing and mailing of—		
American Economic Review	$ 74,508	$ 66,290
Journal of Economic Abstracts	40,870	—
Papers and Proceedings	28,405	23,264
Handbook (September, 1966, supplement)	13,261	—
Provision for *Handbook*	8,000	8,000
Addressing machine rental and expenses	4,979	8,246
Editorial salaries	25,471	20,270
Payments to contributors	1,658	1,387
Editorial supplies and expense	2,921	2,558
Total publication expenses	$200,073	$130,015
ADMINISTRATIVE EXPENSES:		
Secretary's salary, less portion allocated to restricted funds ($6,600 in 1966 and $4,200 in 1965)	$ 8,950	$ 8,367
Office salaries	37,762	33,492
Office machine rentals	656	552
Postage	3,759	3,108
Stationery, printing and supplies	3,050	2,867
Insurance	474	369
Annual meeting expenses	4,700	1,500
Executive Committee expenses	2,052	2,208
Committee on the National Science Foundation Report on the Profession expenses	—	5,162
Committee on Economic Education expenses	—	1,164
Other committees' expenses	1,276	1,346
International Economic Association dues and expenses	600	1,131
Archives	—	3,060
Pension annuity payments	2,128	8,992
Payroll taxes	1,814	1,344
Depreciation of furniture and fixtures	512	448
Telephone and telegraph	1,117	945
Expenses allocated to restricted funds	(3,327)	(477)
Sundry	3,221	2,499
Total administrative expenses	$ 68,744	$ 78,077

REPORT OF THE MANAGING EDITOR
FOR THE YEAR ENDING DECEMBER 1966

The number of manuscripts submitted during 1966 was 451, which is a record high and about double the level of the mid-1950's. Table 1 shows these figures for each year since 1948; it also records the number of manuscripts published each year, and the ratios of published papers to the totals received. During the past year, 62 papers were published and 58 were accepted, reducing a bit the fairly large backlog of papers awaiting publication.

TABLE 1

MANUSCRIPTS SUBMITTED AND PUBLISHED, 1948-66

Year	Submitted	Published	Ratio of Published to Submitted
1948	205	48	.23
1949	200	53	.27
1950	197	53	.27
1951	222	49	.22
1952	190	47	.25
1953	234	51	.22
1954	231	47	.20
1955	245	41	.17
1956	242	48	.20
1957	215	40	.19
1958	242	46	.19
1959	279	48	.17
1960	276	46	.17
1961	305	47	.15
1962	273	46	.17
1963	329	46	.14
1964	431	67	.16
1965	420	59	.14
1966	451	62	.14

Table 2 provides the breakdown of the 1966 volume's contents between articles, review articles, communications, book reviews, etc. Similar information for 1964 and 1965 is also shown. The number of pages devoted to leading articles has increased substantially in the last few years, and the book review section grew in 1966. The decline in listings of periodicals in 1966 was owing to the transfer of this section early in the year to the *Journal of Economic Abstracts*.

TABLE 2

SUMMARY OF CONTENTS, 1964–66

	1966		1965		1964	
	Number	Pages	Number	Pages	Number	Pages
Leading articles..........	25	595	17	443	19	367
Review articles...........	0	0	1	22	2	16
Communications:						
Original...............	13	83	12	110	13	95
Comments and replies..	24	114	29	118	33	131
Books reviewed..........	210	417	189	377	190	386
Classified lists:						
New books............		71		77		76
Periodicals............		15		84		73
Dissertations..........		39		43		47
Notes...................		57		53		46
		1,392*		1,327*		1,237*

* Includes some blank pages.

Table 3 shows the several most popular "fields" in 1966, as reflected in the subject matter of all manuscripts submitted during the year. Almost 25 percent of all manuscripts submitted fell in the three areas of "money," international economics, and economic development. Other popular "fields," not shown in the table, were industrial organization, cost of capital (capital budgeting), education and health, stock market, and unemployment.

TABLE 3

NUMBER OF MANUSCRIPTS SUBMITTED IN THE MOST POPULAR FIELDS IN 1966

Monetary and banking theory and practice...........................	47
International trade and finance......................................	35
Economics of underdeveloped countries..............................	29
Micro-theory..	20
Fiscal and tax theory and policy....................................	19
Production functions; technical change..............................	18

Table 4 summarizes the subject-matter distribution of accepted articles, review articles, and communications for 1966. The most interesting figures are those for articles and original communications, since to some extent these figures indicate the areas in which most work of publishable quality and of broad interest to economists is being done. The table shows, however, that we have accepted manuscripts in almost all of the listed areas.

TABLE 4

Subject-Matter Distribution of Accepted Manuscripts in 1966

	Articles	Review Articles	Original Communi-cations	Comments; Replies	Totals
General economics......		1			1
Price theory............	3		5	5	13
Income theory...........	5		1	2	8
History of economic thought..............					0
Economic history, devel-opment, national eco-nomics..............			1		1
Social accounting........	1		1		2
Economic systems.......			1		1
Business fluctuations.....	1			1	2
Money and banking.....			2		2
Public finance..........	4		1	1	6
International economics..	2			4	6
Business finance........				4	4
Business administration..			1		1
Industrial organization...	1		1	2	4
Land economics; housing.	2			1	3
Labor economics........	1		1		2
Consumption; welfare; population............	2				2
Unclassified............					0
Totals............	22	1	15	20	58

Table 5 presents the actual annual expenditures from 1962 through 1966 for the four regular issues of the *Review* and the estimated budgets for 1966 and 1967. Actual expenditures have risen since 1962 owing principally to increases in printing and mailing expenses and to higher office costs—the former because of the increased number of pages per copy and the larger number of copies printed and the latter because of the appointment of a book review editor and increased expenses associated with the sharp rise in the number of manuscripts handled.

Budgeted expenditures for 1966 were $91,000, while actual expenditures were $100,996. The difference is traced principally to higher printing and mailing costs; my budgeted expenditure for this item was based on an estimate of 86,000 copies of the *Review* for the year and 400 pages per copy, but the actual number of copies was 90,725 and the actual number of pages per copy was 420. Office expenditures were also higher than budgeted by about $1,344.

The 1967 budget shows printing and mailing costs at $82,320, about $7,500 higher than actual costs in 1966. This 1967 figure is based on estimates of 98,000 copies and 420 pages per copy. The budget also shows about a $900 increase over 1966 for office expenses, which reflects a higher salary for my assistant and increased expenses of handling the ever increasing flow of manuscripts, book reviews, and correspondence. The number of manuscripts coming into our office, counting initial submissions, those returned from referees, and authors' revisions, is now more than 850 a year.

TABLE 5
ACTUAL AND BUDGETED EXPENDITURES, 1962–66

	ACTUAL EXPENDITURES					BUDGETS	
	1962	1963	1964	1965	1966	1966	1967
Printing and mailing	$47,785.86	$51,044.09	$53,660.84	$61,606.82†	$74,277.93‡	$65,900.00	$82,320.00‡
Payments to contributors	3,375.00	2,308.00	1,424.50	1,598.00	1,873.50	1,600.00	1,800.00
Subtotal	$51,160.86	$53,952.09	$55,085.34	$63,204.82	$76,151.43	$67,500.00	$84,120.00
Office Expenses							
Editor's salary	$ 8,000.00	$ 8,000.00	$ 8,000.00	$ 9,000.00	$ 9,000.00	$ 9,000.00	$ 9,000.00
Editorial assistance	11,885.05	9,702.06	13,460.06	12,024.08	14,189.83	12,900.00	15,125.00
Assistant to editor	8,100.00*	6,234.44	6,619.32	6,619.32	7,330.00	7,000.00	8,025.00
Book reviews	—	—	1,833.32	2,000.00	2,000.00	2,000.00	2,000.00
Secretarial	1,664.38	1,442.25	1,632.75	1,725.00	2,160.15	1,800.00	2,400.00
Clerical, proofreading, etc.	2,120.67	2,025.37	3,374.67	1,679.76	2,699.68	2,100.00	2,700.00
Postage and supplies	851.95	1,165.50	1,348.23	1,462.32	1,479.61	1,500.00	1,500.00
Office equipment	—	812.63	122.07	—	175.50	100.00	100.00
Subtotal	$20,737.00	$19,680.19	$22,930.36	$22,486.40	$24,844.94	$23,500.00	$25,725.00
Total	$71,897.86	$73,632.28	$78,015.70	$85,691.22	$100,996.37	$91,000.00	$109,845.00

* Includes $1,500 to acting Managing Editor.
† Corrected figure.
‡ Based on estimated copies of 98,000 and 420 pages per copy.

Table 6 gives detailed information about printing and mailing costs by is-
sues during 1966. The number of copies printed in 1966 averaged 22,681 per
issue compared to 20,400 the year before and to 18,500 in 1964. Ten years
ago, the figure was only 11,550. The successive increases are of course mainly
related to the increases in memberships and subscriptions.

TABLE 6

COPIES PRINTED, SIZE, AND COST OF PRINTING AND MAILING IN 1966

	PRINTED	PAGES		COST		
		Net	Gross	Issue	Reprints	Total
March........	22,500	331	416	$17,840.79	$122.50	$17,963.29
June.........	22,225	346	400	17,456.08	118.08	17,574.16
September....	23,000	384	448	20,130.79	113.31	20,244.10
December....	23,000	331	416	18,388.00	108.38	18,496.38
	90,725	1,392	1,680	$73,815.66	$462.27	$74,277.93

During the year I have had the advice and assistance of the following for-
eign correspondents, who have been particularly helpful with regard to the se-
lection of foreign books for listing and review:

Maurice Flamant (France)
Erich Schneider (Germany)

Two members of the Board of Editors complete their three-year terms of
office at this time: Richard Caves and Arthur Goldberger. The Association
owes them a heavy debt of gratitude for the generous expenditure of time they
have made in the interests of the *Review*, and I very much appreciate their
constant willingness, even when other obligations pressed, to review manu-
scripts and offer editorial advice. I nominate as their successors on the Board,
for three-year terms beginning in 1967: Ronald Winthrop Jones and Dale W.
Jorgenson.

In June, my assistant, Jane Ladd, resigned to accept another position. This
unfortunate event, however, was soon followed by a happy one when Doris
Merriam agreed to return to the fold. Miss Merriam was the assistant to two
managing editors from 1946 to 1962.

During the year I have frequently sought the aid of members of the profes-
sion in addition to the members of the Editorial Board—partly to relieve the
latter of what would otherwise be an impossibly heavy burden and partly to
obtain advice of specialists in particular areas not represented on the Board.
The following have assisted in this way:

M. Abramovitz	K. Arrow	G. Becker	W. Bowen
M. Adelman	H. Averch	A. Bergson	H. Brazer
W. Alberts	G. Bach	J. Bhagwati	G. Break
A. Alchian	J. Bain	M. Blaug	H. Brems
E. Ames	R. Baldwin	G. Borts	M. Brennan
W. Anderson	D. Bear	K. Boulding	E. Brigham

M. Bronfen-
brenner
M. Brown
O. Brownlee
J. Chipman
R. Clower
R. Coen
P. Cootner
M. David
P. David
O. Davis
H. Demsetz
T. Dernburg
E. Despres
P. Diamond
C. Díaz-Alejandro
E. Domar
H. Dougall
O. Eckstein
W. Fellner
C. Ferguson
F. Fisher
B. Fleisher
K. Fox
M. Friedman
I. Friend
E. Furubotn
L. Gallaway
F. Gehrels
D. Gordon
M. Gordon
R. Gordon
M. Gort
P. Gray
H. Grubel

J. Guttentag
F. Hahn
B. Haley
C. Hall
J. Harsanyi
B. Hickman
W. Hochwald
Y. Ijiri
J. Ingram
F. Johnston
D. Jorgenson
A. Kahn
E. Kane
J. Kareken
P. Kenen
C. Kennedy
F. Kidner
C. Kindleberger
A. Klevorick
L. Krause
M. Kreinin
R. Kuenne
J. Kuhlman
M. Kurz
L. Lave
H. Leibenstein
E. Lerner
A. Levenson
H. Levine
H. Liebhafsky
M. Lovell
F. Machlup
G. Maddala
S. Maisel
B. Malkiel

S. Marglin
J. Margolis
A. Marty
T. Mayer
S. McDonald
R. McKean
L. McKenzie
J. McKie
R. McKinnon
A. Meltzer
M. Michaely
H. Miller
M. Miller
E. Mills
J. Mincer
F. Modigliani
R. Moorsteen
J. Morgan
L. Moses
R. Murphy
G. Nutter
W. Owen
M. Paglin
S. Peltzman
R. Pfouts
E. Phelps
R. Piron
J. Porterfield
R. Powell
R. Quandt
M. Reder
A. Rees
L. Reynolds
R. Rhomberg
A. Rivlin

E. Rolph
G. Rosenbluth
J. Rosse
J. Rothenberg
S. Rottenberg
W. Salant
A. Sametz
P. Samuelson
A. Scott
R. Selden
E. Shaw
C. Shoup
N. Simler
H. Simon
R. Sliter
W. Smith
E. Solomon
G. Somers
B. Stein
J. Stein
G. Stigler
D. Suits
R. Summers
P. Taubman
R. Teigen
D. Tucker
L. Ulman
W. Vickery
J. Walter
W. White
O. Williamson
D. Winch
P. Wonnacott
L. Yeager

Respectfully submitted,
JOHN G. GURLEY, *Managing Editor*

REPORT OF THE COMMITTEE ON RESEARCH AND PUBLICATIONS

In the "Reading Series" the volume on business cycles, edited by R. A. Gordon and Lawrence R. Klein, was published. A volume on international economics, edited by Harry G. Johnson and Richard Caves, was sent to the publishers. The volume on welfare economics, edited by Kenneth Arrow and Tibor Scitovsky, should be sent to the publishers in the very near future. Work is progressing on a book of readings on agricultural economics, edited by Gale Johnson and Karl Fox; and Robert Solow and W. Arthur Lewis have agreed to edit a new volume on growth. The Committee is investigating means of making the readings volumes available more cheaply.

The first volume in the series of translations of foreign economic works has been sent to the publisher. It is the *Theory of the Peasant Economy,* by Chayanov. Other translations in process include Tugan-Baranovsky, *The Russian Factory;* Dahmen, *Svensk Industriell Företagarverkamnet;* Pareto, *Manuale;* Galiani, *Della Moneta.* In additon, essays from the economic history of China, France, and Japan are in various stages of planning or completion. The Committee invites further suggestions from the profession concerning important foreign works that ought to be translated into English, both from those who would like to see such translations available and from those who would be willing to prepare them. Professor Bert F. Hoselitz, of the University of Chicago, is responsible for this aspect of the Committee's work.

A series on policy-oriented surveys of economics in foreign countries is being prepared under the editorship of Harry G. Johnson. Surveys for eight countries have been arranged and manuscripts are expected by the end of 1967. It is hoped to publish these as supplements to the *American Economic Review.* It is planned also to publish this series as a volume, perhaps in conjunction with the earlier series of surveys of foreign economics under the editorship of George Hildebrand. The volume including the Hildebrand series would contain a survey of Italian economics by Professor Lombardinni received too late for incorporation in the series as originally published.

IRVING B. KRAVIS, *Chairman*

REPORT OF COMMITTEE ON ECONOMIC EDUCATION

During 1966, the Committee on Economic Education has continued to devote its energies primarily to development of a more active program in the area of economic education, at college and pre-college levels, in accordance with the policy statement adopted by the Executive Committee in March, 1964. Six lines of activity are worthy of mention.

1. Early this year, the Committee arranged a grant from the National Science Foundation to the Association to inaugurate a "Visiting Scientist Program in Economics," comparable to those in a number of other sciences. This program was announced to the profession last spring and went into effect with the 1966-67 academic year. It provides for visits of about twenty distinguished economists to smaller campuses, generally those concentrating on undergraduate education, to stimulate student interest in further work in economics, to give seminars, to consult with members of the host department, and, where desired, to consult with college officials concerning programs in economics. An informal roster of some seventy-five leading economists has been established to provide visitors under this program. We are currently working to assure continuation of the program in the following academic year.

2. Last year this Committee recommended to the Joint Council on Economic Education, with which we have been working closely, establishment of a new committee of well-known economists, representing a cross-section of the profession, and leading psychologists to prepare a widely acceptable "Test of Understanding in College Level Economics," which might be used especially as a research and evaluative instrument for departments wanting to analyze the effectiveness of their basic courses in economics. With financing arranged by the Joint Council, a special *ad hoc* committee was established, under the chairmanship of Professor Rendigs Fels, and the work of his committee is now well under way. Professor Fels's report on the committee's work to date is included in this volume. The tests should be ready for national use by September, 1967.

3. Last year, again working with the Joint Council on Economic Education, this Committee helped arrange a special grant from the Kazanjian Foundation to finance publication of reports on interesting experiments in the teaching of basic economics in the colleges and universities. Departments throughout the country were invited to submit such experiments, together with information on their evaluation. Some thirty responses were received by the Committee, and Professor Bernard Haley was invited to prepare an analytical summary of these experiments, plus any others that he might obtain information on. Professor Haley's analysis is complete. A summary is included in this volume, and a more complete report is being released currently through the Joint Council on Economic Education. Copies can be obtained from the Joint Council.

4. Three members of this Committee serve as a special advisory committee to the Joint Council on Economic Education, and informally the entire Committee so serves. Under this arrangement, we have worked closely with the

Joint Council in establishing an extensive publications program, involving preparation of special teaching materials in economics for use in high school courses in American history, problems of democracy, civics, and the like. Small "units" involving economic concepts and economic analysis are being prepared by well-known economists, working with high school educators, in some ten areas, mainly for use in the ninth through twelfth grades. Drafts of some of these units are now being tried out in high schools; others are scheduled for early completion. After testing and revision, all units will be released for general use through cooperative arrangements between the Joint Council and commercial publishers.

5. The Joint Council, with the advice and consultation of this Committee, is establishing a new college-university program which is expected to include sponsorship of a large-scale research project to investigate the lasting effects of college and university basic economics courses. It is also expected to include special grants to a small number of departments of economics throughout the country to help finance experimentation in the teaching of basic economics and careful evaluation of the results of such experiments.

6. During the past two years, members of this Committee have been increasingly involved in attempts to represent the profession in contacts with government agencies, especially the United States Office of Education and the National Science Foundation, but also state and local educational agencies. Requests have come for assistance in planning new courses, preparing and selecting materials, and the like, at state and local levels. At the national level, the large expansion of funds being poured into education and the increasing inclusion of economics as one of the areas for support at the grade and high school levels have involved a mushrooming amount of planning activities and committees to establish procedures, allocate funds, and the like. Unless economics is represented in such discussions, as a practical matter the funds tend to go elsewhere, for teacher workshops, for research undertakings, for teaching experiments, and the like. Many other professions have either permanent secretaries in Washington or part-time employees carrying on these activities. Members of this Committee, with the active participation of the Secretary of the Association, have put in a substantial amount of time on such activities. We suspect the profession may soon face the need to establish more formal arrangements if it wants to be represented actively and adequately in such activities vis-à-vis the federal government and other governmental agencies.

G. L. BACH, *Chairman*

REPORT OF COMMITTEE ON FINANCIAL AND
INVESTMENT POLICIES

1. *Present method of reporting income from equities:* Currently, income from equities is reported as the sum of dividends received and capital gains realized.

2. *Defects of present method:* This method has three main defects. First, it takes no regular account of earnings of enterprises plowed back rather than paid out as dividends. Second, it does not recognize the market's evaluation of changes in the future prospects of the enterprises whose stocks we own. Third, it makes income depend on the accident of when securities are sold. The first leads to an understatement of income; the second, in principle might lead to either understatement or overstatement, but in practice has led to understatement. The third produces wide fluctuations in recorded income that do not necessarily correspond to changes in the amount available for expenditures without impinging on capital.

3. *Proposed alternative:* We recommend remedying all three defects as follows:

(3.1) *Retained earnings:* Instead of recording as income simply dividends received, we propose including also retained earnings. The easiest way to do this as an accounting matter is to record as income the earnings per share for each company whose shares we own times the number of shares owned. For shares purchased or sold during the year, we propose including one-half of the annual earnings. The accounting procedure would be to recognize dividends as income when received, and at the end of the period to enter as an additional item of income the excess of earnings as calculated above over dividends received. For firms whose fiscal year does not coincide with the Association's fiscal year, the earnings to be included would be those for the corporation's fiscal year ending within the Association's fiscal year. (See paragraph 7.)

(3.2) *Changes in prospects of enterprises:* To eliminate the second effect, we propose computing for each year the changes in the market value of the stocks in our portfolio adjusted for (*a*) retained earnings and (*b*) changes in the general price level. Adjustment (*a*) will be made by subtracting retained earnings as calculated above. Adjustment (*b*) will be made by subtracting the product of the start of the year market values of the portfolio times the percentage change in a price index. We propose to use as the price index the implicit GNP deflator, and to measure the percentage change from fourth quarter to fourth quarter. However, the adjusted change in market value computed in this way will not be entered in full each year in order to overcome the third defect discussed in the next paragraph.

(3.3) *Realized capital gains and associated arbitrary fluctuations:* By recognizing as in the previous paragraph, all changes in market value, we would remove the fortuitous element in our present method arising from the timing of sales. But, including all such gains would give undue significance to short-term market fluctuations. Since the objective is to take account of longer-term

movements in stock prices, we propose spreading the gain as calculated in the previous paragraph over three years, including one-third in the year to which the calculation refers and one-third in each of the next two years. The number "three" was selected as roughly corresponding to the duration of cycles in stock market prices in the post-World War II period.[1]

(3.4) *Final formula:* This treatment would make the Association's income from equities in each year equal to dividends received in that year plus our share of earnings plowed back that year plus one-third of market appreciation in that year (adjusted for retained earnings and price-level change) plus one-third of adjusted appreciation in the preceding year, plus one-third of adjusted appreciation in the second preceding year.

4. *Comparison of results under various methods:* Table 1 shows what the Association's reported income would have been from 1955 to 1965 if it had been calculated by a number of alternative methods. To simplify the calculations it was assumed that the cost of the 1952 year end portfolio was identical with its then existing market value. The methods compared are: (1) The present method (dividends plus realized gains); (2) dividends plus retained earnings; (3) dividends plus total change in market value; (4) dividends plus total change in market value less adjustment for price-level change; (5) the proposed alternative (dividends plus retained earnings plus three-year average of adjusted change in market value). Table 1, columns 1 to 5 give dollar income; columns 6 to 10 give rates of return, calculated as a percentage of the average market value of the portfolio during the year. Table 2 gives the basic data used in calculating Table 1. Our present method has understated income substantially and reported it as varying widely; the proposed method recognizes income fully and as varying much less.

5. *Transition:* We propose this procedure take effect for the 1967 fiscal year. At that time the Members' Equity and the reported income for the two prior years would be restated on the new basis.

6. *Auditor's participation:* We wish to express appreciation to Arthur Andersen & Co. for their cooperation and counsel in preparing the data of Tables 1 and 2 and in discussing the problem of how to account for income from equities. They have expressed the informal opinion that the method of reporting income proposed is in conformity with generally accepted accounting principles.

7. *Fiscal year:* The Association's fiscal year now ends November 30. The reason has been to have financial accounts available for the December meetings. Since most corporations prepare their accounts on a calendar basis, a fiscal year ending December 31 would fit the proposed method much better. We propose a change to a calendar-year basis; the financial accounts could then be available for the spring meeting of the Executive Committee.

MILTON FRIEDMAN, *Chairman*

[1] In the post-World War II period, peaks in stock prices were reached in 1946, 1948, 1953, 1956, 1959, 1961 and 1966. There were six cycles in the 20 year period, or an average duration of 3⅓ years. In the post-World War I period, 1919 to 1939, there were only six peaks, or an average duration of 4 years.

TABLE 1

	ANNUAL INCOME					RATES OF RETURN				
	1	2	3	4	5	6	7	8	9	10
	Present Method (Dividends Plus Realized Gains)	Dividends Plus Retained Earnings	Dividends Plus Total Change in Market Value	Dividends Plus Total Change in Market Value Less Adjustment for Price Level Change	Proposed Alternative (Dividends Plus Retained Earnings Plus Three-Year Average of Adjusted Change in Market Value)	Method 1	Method 2	Method 3	Method 4	Method 5
1953	2,048	7,655	5,234	4,422		2.63	9.85	6.73	5.69	
1954	8,089	5,919	19,000	18,027		11.81	8.64	27.45	26.59	
1955	8,538	6,555	21,858	20,832	14,272	10.40	7.98	26.61	25.37	17.38
1956	3,127	7,591	5,656	2,465	14,677	3.29	7.98	5.95	2.59	15.43
1957	19,471	6,856	7,692	4,098	8,987	22.28	7.84	8.80	4.69	10.28
1958	12,163	6,001	25,129	23,128	9,082	13.42	6.62	27.73	25.52	10.02
1959	10,952	7,121	29,091	27,442	18,685	9.77	6.35	25.95	24.48	16.67
1960	16,455	7,596	9,768	7,745	20,129	12.52	5.78	7.43	5.89	15.32
1961	16,205	8,615	49,462	47,676	28,459	9.53	5.07	29.10	28.05	16.74
1962	4,912	9,432	−22,577	−24,554	11,143	2.75	5.29	−12.66	−13.77	6.25
1963	7,363	10,730	31,390	29,168	18,567	4.05	5.90	17.27	16.05	10.21
1964	15,043	12,644	32,320	29,041	12,927	7.18	6.04	15.43	13.87	6.17
1965	8,634	15,901	46,238	42,124	36,254	3.38	6.22	18.08	16.47	14.18
Mean 1955–65	11,169	9,004	21,457	19,015	17,562	8.96	6.46	15.43	13.56	12.60
σ 1955–65						5.85	1.04	12.53	12.77	4.17

Column 1 = Table 2, Column 1 + Table 2, Column 7
Column 2 = Table 2, Column 1 + Table 2, Column 2
Column 3 = Table 2, Column 1 + Table 2, Column 4
Column 4 = Table 2, Column 1 + Table 2, Column 4 minus Column 3
Column 5 = Table 2, Column 1 + Table 2, Column 2 + Table 2, Column 6

Column 6 = Table 1, Column 1 + Table 2, Column 10
Column 7 = Table 1, Column 2 + Table 2, Column 10
Column 8 = Table 1, Column 3 + Table 2, Column 10
Column 9 = Table 1, Column 4 + Table 2, Column 10
Column 10 = Table 1, Column 5 + Table 2, Column 10

TABLE 2

BASIC DATA

	1	2	3	4	5	6	7	8	9	10
	Dividends	Retained Earnings	Price Level Change in Opening Market Value	Change in Market Value	Change in Market Value Adjusted for Price Level Change and Retained Earnings	Three-Year Average Change in Market Value Adjusted for Price Level Change and Retained Earnings	Realized Gains		Market Value 11/30/$t-1$	Average Market Value Year t
1953	4,185	3,470	812	1,049	−3,233		−2,137		89,269	77,732
1954	2,552	3,367	973	16,448	12,108		5,537		66,196	68,480
1955	2,486	4,069	1,026	19,372	14,277	7,717	6,052		70,766	82,127
1956	3,105	4,486	3,191	2,551	−5,126	7,086	22		93,589	95,107
1957	2,898	3,958	3,594	4,794	−2,758	2,131	16,573		96,626	87,402
1958	3,023	2,978	2,001	22,106	17,127	3,081	9,140		78,179	90,633
1959	3,469	3,652	1,649	25,622	20,321	11,563	7,483		103,089	112,112
1960	4,063	3,533	2,023	5,705	149	12,532	12,392		121,136	131,422
1961	4,394	4,221	1,786	45,058	39,061	19,844	11,811		141,708	169,979
1962	4,735	4,697	1,977	−27,312	−33,986	1,741	177		188,251	178,300
1963	5,091	5,639	2,222	26,299	18,438	7,834	2,272		168,349	181,775
1964	5,421	7,223	3,279	26,899	16,397	283	9,622		195,200	209,403
1965	6,441	9,460	4,114	39,797	26,223	20,353	2,193		223,606	255,702
1966									287,799	
Mean	4,102	4,901	2,442	17,355	10,011	8,560	7,067			

REPORT OF THE JOINT COMMITTEE OF THE AMERICAN ECONOMIC ASSOCIATION AND THE ASSOCIATION OF AMERICAN LAW SCHOOLS

The *ad hoc* Joint Committee of the A.E.A. and the A.A.L.S. met in Chicago on Saturday, November 12, 1966. Eight of the ten members of the Committee were present. There was a general discussion of the contributions law and economics could make to each other and various devices which could be used to improve the dialogue between lawyers and economists were considered.

It was the consensus of the group that law professors today frequently write about legal subjects without sufficient awareness of the economic implications of their statements. In considerations of policy in many areas, for example, a greater use of economic analysis would frequently make this legal work more cogent, thus leading to the formulation of more desirable policies. In addition to correction of implicit and explicit errors of economic analysis, legal writing could be improved by new modes of formulating issues which economics offers. Finally, economics may be able to improve academic legal analysis by suggesting those areas in which empirical proof for or against a proposition is available or is needed.

In considering what law has to offer academic economists, it was recognized that many economists are unfamiliar with the process by which legal developments occur and, therefore, fail to understand why most legal literature does not proceed like an article in an economics journal. But the principal problem was conceived to be more serious; that is, that much of contemporary economics literature simply lacks relevancy for current problems. The lawyer is at the forefront of these issues; in fact, he frequently is instrumental in instituting change. Law journal literature is generally more topical than that in economics journals. Thus there exists a peculiarly undesirable time lag in the interacting of law and economics. Too frequently social or economic developments initiated, or at least closely followed, by lawyers are over and done with before the economist begins to consider them. It was the feeling of the group that the entire process of policy formulation and implementation could be improved if economists dealt with problem areas as they arose rather than from the more comfortable but less helpful vantage point of hindsight.

It was also felt that law and the legal literature present many areas of challenge and interest to economists but of which they are unaware. Thus, areas of well-established law as well as those just developing should be studied and analyzed by economists as well as lawyers. Finally, it was recognized that there are legal areas in which both lawyers and economists have traditionally worked but often at cross-purposes or without adequate communication between the two groups.

Thus the Committee felt that any proposal it made should serve the dual purposes suggested above. The lawyers in a field should be informed about what economics has to offer, both in an affirmative and a critical sense; and economists should be informed about current areas requiring their expertise,

as well as more traditional areas where study in depth could be helpful. Several methods for meeting these goals were considered, but one plan seemed clearly to have advantages over any other. That plan, which is hereby proposed by your Committee, would operate substantially as follows: A competent economics scholar would be paid to prepare a monograph on the legal literature of a specified area of law. This monograph would constitute a critique of the economic content of the legal literature. Errors in economics as well as omissions of relevant economic inquiry would be pointed out and proposals for improving the formulation of policy positions would be made. This monograph would then be circulated to interested parties and a conference on the subject would be held.

It seemed appropriate to your Committee to try this proposal on a limited experimental basis before proceeding to urge its broader use. Consequently, three areas were selected as posing different kinds of problems, one which is currently in the process of new development, one which has recently seen a major shift in the law's position, and one in which both lawyers and economists have worked for some time.

For the first of these, we selected the area of criminal law. Here it was recognized that crime represents a cost to victims and a cost to the public for enforcement of the criminal law. But there are many variables, including crime prevention, detection, adjudication, punishment, and rehabilitation. Many recommendations are made by lawyers for more or less expenditures on one or more of these factors. This involves problems of techniques familiar to economists, but generally the legal literature has failed to take adequate note of the contribution economic theory offers. On the other hand, this area of very sensitive social policy has received scant attention from economists, though one or two are known presently to be working in the area.

The second legal area for consideration is that of product liability (perhaps also including safety legislation and automobile accident compensation). This area of law is undergoing a substantial and radical change, generally in the direction of shifting a large part of the cost of faulty goods back to the producer rather than leaving it with the consumer. This development occurred almost entirely within the courts, and the relevant literature appears in the law reviews. It is not known that any economists have worked on this problem per se, although there are a number of related areas in which substantial work has been done.

The third area is antitrust. Here, of course, the implications of economics for legal analysis are very evident, and there has indeed been considerable dialogue between lawyers and economists on the subject. Nonetheless, it was deemed desirable to formulate the same kind of monograph for this field as for the others, since no single critical appraisal of the legal literature exists. It was felt, however, that this area was too large to give to one person, and consequently the committee recommends that this assignment be divided among two individuals.

After the Committee has made a selection of an individual to prepare each monograph, he would be given approximately one summer in which to complete most of his work. The Committee will also provide him with either the

names of law professors in the area or an actual reading list, but the Committee felt that a collaborative effort was not indicated, since the desired critical flavor might thereby be lost. The economist can, of course, consult with anyone as much as he wishes, but the responsibility for the monograph should be solely his.

On completion of the monograph it would be distributed for comments to a prominent group of readers in both law and economics for their appraisal. Then a conference with approximately twenty invited participants, half lawyers and half economists, would be held to discuss the monograph and the appraisals. Arrangements should definitely be made for publication of the monograph, and there is a possibility that publication arrangements could be made for the proceedings of the conferences.

The total budget for this first effort should be approximately $40,000, distributed as follows: 4 monographs at $5,000, $20,000; conference expenses, $15,000; publication and miscellaneous expenses, $5,000.

There are a number of foundations or other agencies who might be approached for the necessary financing. These include the American Bar Foundation, Walter E. Meyer Foundation, Russell Sage Foundation, Ford Foundation, Social Science Research Council, Carnegie Foundation, National Institute of Health (for product liability and safety area), and National Science Foundation (for criminal law area).

Therefore, it is respectfully requested by your Committee that permission be given by the Executive Committee of the Association of American Law Schools and by the Executive Committee of the American Economic Association to proceed with efforts to raise money for this project. If such approval is given, it is also recommended by your Committee that this Joint Committee be established as a permanent standing committee of the two associations.

For the A.E.A.:
ARMEN A. ALCHIAN
RONALD H. COASE
GERALD M. MEIER
GEORGE J. STIGLER

For the A.A.L.S.:
HARLAN M. BLAKE
ROBERT H. BORK
WILLIAM A. KLEIN
JAMES A. RAHL
HENRY G. MANNE, *Chairman*

REPORT OF REPRESENTATIVE TO THE
NATIONAL BUREAU OF ECONOMIC RESEARCH

History and Organization of National Bureau. The National Bureau was organized in 1920 in response to a growing demand for objective determination of the facts bearing upon economic problems and for their interpretation in an impartial manner. It is a private, nonprofit organization. Its Board of Directors includes men of different economic and social viewpoints, chosen from widely differing spheres of activity. Members include persons from industry and labor; banking and finance; professional associations of economists, economic historians, accountants, and statisticians; and economists from university faculties. The American Economic Association is one of the representative organizations each of which nominates a member of the National Bureau's Board of Directors.

Staff and Conference Programs. Staff research in 1966 included investigations in the following areas: national income and wealth; economic growth; business cycles; prices, costs, and profits; labor markets; money, banking, and finance; governmental activity; and international economic relations. Among new studies started during the year were those on the structure of national accounts by Richard and Nancy Ruggles, the economics of health under Victor R. Fuchs, and banking structure and performance under Donald P. Jacobs.

The conference program consisted of a conference sponsored by the Universities-National Bureau Committee for Economic Research on the Economics of Defense, April, 1966, and two conferences sponsored by the Conference on Research in Income and Wealth, one on Review of the United Nations Proposals to Revise Their System of National Accounts, June, 1966, and the other on Industrial Composition of Income and Product, December, 1966.

Fourteen reports resulting from the National Bureau's program were published in 1966 and ten others were in press in December. Of these twenty-four publications, eight are books by the staff, ten are Occasional Papers by the staff, and six are books of conference proceedings.

At the annual meeting of the Universities-National Bureau Committee for Economic Research in April, 1966, it was voted to continue the Committee for another four-year term, July 1, 1966-June 30, 1970 and to add the following universities to the Committee: Ohio State University, University of Pittsburgh, and Washington University (St. Louis).

Meetings of the Board of Directors in 1966. The 47th Annual Meeting of the National Bureau was held Monday, March 7, 1966, and the National Bureau's 46th Annual Report, "Anticipating the Nation's Needs for Economic Knowledge," was presented to the Board. Copies of the report are available upon request to the National Bureau.

A fall meeting of the Board was held November 3-5 to review the current and planned research program.

Directors, Officers, and Staff. At the 1966 Annual Meeting of the National

Bureau Joseph A. Beirne and Robert V. Roosa were elected Directors at large. The American Finance Association was designated an additional representative organization to have a member on the National Bureau's Board, and Walter E. Hoadley was subsequently elected to this position.

Officers elected were Frank W. Fetter, Chairman; Arthur F. Burns, President; Theodore O. Yntema, Vice President; Donald B. Woodward, Treasurer; Geoffrey H. Moore, Director of Research; Douglas H. Eldridge, Executive Director and Secretary; Hal B. Lary and Victor R. Fuchs, Associate Directors of Research.

William J. Carson resigned as Secretary of the Board at the end of February, 1966.

Sam Peltzman, University of California, Los Angeles, was appointed a research fellow for 1966-67.

William H. Brown, Jr., member of the research staff since 1960, died on January 20, 1966. N. I. Stone, Director Emeritus and a member of the Bureau's first Board of Directors, died on October 24, 1966.

WILLARD L. THORP

REPORT OF THE CENSUS ADVISORY COMMITTEE

The Census Advisory Committee of the American Economic Association met with the Director and Staff of the Census Bureau in January, 1966. The Deputy Assistant Secretary of Commerce and a representative of the Office of Business Economics also joined in the meeting.

The Committee reviewed the Bureau's plans for the 1970 Census of Population and other existing programs and offered suggestions for their improvement designed to reflect current and future data needs of the country's economic problems.

In response to a recommendation by the Chairman of the Committee at an earlier meeting, the Bureau reported its decision to publish a report on the industrial composition of the population, based on the 1960 Census of Population. This report, which would correspond to a similar report published for the 1950 Census of Population, is scheduled to be released in early spring of 1967.

The Bureau staff described to the Committee their plans for the 1967 Economic Censuses to substitute Internal Revenue Service tax return computer records for over one million small employer firms formerly covered in the Census mail canvass.

Committee members—and through the Committee, members of the American Economic Association—were informed on the Bureau's continuing efforts to make Census data more accessible to users outside the government, such as making available for sale various data tapes and computer programs. The scope of the Bureau's publication catalogue has been expanded to include a section on unpublished materials as well as a section on publications.

The Committee also reviewed and commented on a second draft of the Bureau's proposed annual report on long-term economic growth. Many of the revisions contained in this version of the report were the result of suggestions made by Committee members at previous meetings. This report, which was subsequently published in October, 1966, under the title of *Long Term Economic Growth,* is a statistical compendium developed from many government and nongovernment sources to provide a new research tool for the study of long-term trends in the U.S. economy.

SOLOMON FABRICANT, *Chairman*

REPORT OF REPRESENTATIVE TO BEHAVIORAL
SCIENCES DIVISION, NATIONAL RESEARCH COUNCIL

Several changes in personnel have taken place. A new Chairman has assumed office, Kingsley Davis, professor of sociology at Berkeley. The Chairman designate is Herbert Simon, who represents the Social Science Research Council. A new Executive Secretary has assumed office, Henry David, formerly professor of economics and president of the New School for Social Research. He has been designated as Pitt Professor of American History and Institutions at Cambridge in 1969-70. I believe this leadership will be effective and promises a larger role for social science and in particular for economics in the Division. I am happy that James Tobin has accepted membership of the Division as a representative of the American Economic Association. I hope that similarly outstanding appointments will continue to be made by our presidents.

Liaison members of the Division have not been appointed from government agencies. However, Howard Hines, who is director of the Social Sciences Division of N.S.F., was present as a guest at the annual meeting. I believe his participation in informal discussions was not without effect. Also, no more societies have been invited to name representatives, although additional members-at-large may be proposed from related disciplines in the future.

There are now several committees in operation which should interest economists.

1. A committee has been established under the chairmanship of Ernest R. Hilgard from psychology to prepare Planning Reports for the Behavioral Sciences in which their need for government support will be detailed. Carl Kaysen is the member for economics and others may be named. Henry Riecken, of the S.S.R.C., is Vice-Chairman. The report of the Committee will be reviewed both by the S.S.R.C. and by the Committee on Public Policy (COS-PUP) of the National Research Council. The final report will be transmitted to the Governing Board of the N.R.C. by the Division. However, the Division will exercise no supervision over the Committee, nor can it reject the report. It was felt that the participation of S.S.R.C. in this project was an assurance that the interests of social science would be observed. The Committee's report is due in 1968. In view of the recent congressional interest in supporting social science, this report could have a major impact on the support of our research. The Committee is financed by a grant from N.S.F.

2. A committee is now active on Government Programs in the Behavioral Sciences. The Chairman is Donald Young, of Rockefeller University, and the Associate Chairman is Herbert Simon. The Executive Secretary is Gene Lyons, of the Public Affairs Center, Dartmouth College. He will also prepare a monograph on government-sponsored research. The only economist on the Committee at present (*pace* Herbert Simon) is Thomas Schelling. It seems clear that the committee intends to discuss all relations between government and social science research, and not merely foreign area research. They have, for example, taken an interest in the various hearings and bills before congres-

sional committees that have recently occurred on possible extensions of support for social science. Because of this I have pressed for a wider representation from economics. I do not know that anything has been done though some people spoke favorably of the idea. The Committee is financed by ONR and the Russell Sage Foundation.

3. A Committee has been established to consider International Relations in the Behavioral Sciences. Evidently not all social sciences have international societies which are as effective as the International Economic Association and the Econometric Society. Lawrence Klein is a member of this Committee for economics. The Chairman is Martin Lipset, who is professor of sociology at Harvard.

4. A Committee has been working for some time under the chairmanship of David Easton on Information in the Behavioral Sciences. Its activities are now supported by N.S.F. It is concerned with the collection and processing of data and with the creation of information networks. Evidently the Committee's performance is well liked by those familiar with its history. A final report is being prepared.

5. A new committee has been proposed by Kingsley Davis to consider the Social Aspects of Science. It would inquire into the conditions and requirements for the development of science.

My general impression is that the Division is moving in the right direction. However, a complete reformation is not to be expected overnight. It is difficult for the social sciences to achieve full recognition in an organization which has been dominated by the physical and biological sciences from its creation. Of course the structure of the N.R.C. reflects the structure of the National Academy of Science and the National Science Foundation. In a worldwide perspective it is unusual for social scientists to be omitted from membership in the National Academy. The question is naturally posed whether the social sciences should seek their own foundation, as some in Congress propose, and perhaps their own academy, as the engineers have done, or should seek the shelter and support of organizations where the "harder" sciences predominate. I think opinions (and feelings!) are divided on this issue in the Division and I gather in the profession at large. Economists should consider the views they will wish to push, since events may move rapidly in the near future.

LIONEL W. McKENZIE

PUBLICATIONS

OF THE

AMERICAN ECONOMIC ASSOCIATION

1967

FIRST SERIES

The supply of those double starred is exhausted. For information apply to the Secretary.

Volume I, 1886

Price in paper

1. **Report of Organization of the American Economic Association, Pp. 46. $ .50
2-3. **Relation of the Modern Municipality to the Gas Supply. By E. J. James. Pp. 66. .75
4. Co-operation in a Western City. By Albert Shaw. Pp. 106. .75
5. **Co-operation in New England. By E. W. Bemis. Pp. 136. .75
6. **Relation of the State to Industrial Action. By H. C. Adams. Pp. 85. .75

Volume II, 1887

1. **Three Phases of Co-operation in the West. By Amos G. Warner, Pp. 119. .75
2. **Historical Sketch of the Finance of Pennsylvania. By T. K. Worthington. Pp. 106. .75
3. **The Railway Question. By Edmund J. James, Pp. 68. .75
4. **Early History of the English Woolen Industry. By W. J. Ashley. Pp. 85. .75
5. **Mediaeval Guilds of England. By E. R. A. Seligman. Pp. 113. .75
6. **Relation of Modern Municipalities to Quasi-public Works. By H. C. Adams and others. Pp. 87. .75

Volume III, 1888

1. Statistics in College, by C. D. Wright; Sociology and Political Economy, by F. H. Giddings; The Legal-Tender Decisions, by E. J. James. Pp. 80. .75
2. **Capital and Its Earnings. By John B. Clark. Pp. 69. .75
3. The Manual Laboring Class, by F. A. Walker; Mine Labor in the Hocking Valley, by E. W. Bemis; Report of the Second Annual Meeting. Pp. 86. .75
4-5. **Statistics and Economics. By Richmond Mayo-Smith. Pp. 127. 1.00
6. **The Stability of Prices. By Simon N. Patten. Pp. 64. .75

Volume IV, 1889

1. **Contributions to the Wages Question: The Theory of Wages, by Stuart Wood; Possibility of a Scientific Law of Wages, by J. B. Clark. Pp. 69. .75
2. **Socialism in England. By Sidney Webb. Pp. 73. .75
3. Road Legislation for the American State. By J. W. Jenks. Pp. 83. .75
4. Third Annual Meeting: Report of the Proceedings. Pp. 123. .75
5. **Malthus and Ricardo, by S. N. Patten; The Study of Statistics, by D. R. Dewey; Analysis in Political Economy, by W. W. Folwell. Pp. 69. .75
6. **An Honest Dollar. By E. Benjamin Andrews. Pp. 50. .50

Volume V, 1890

1. **The Industrial Transition in Japan. By Yeijro Ono. Pp. 122. $1.00
2. Two Essays on Child-Labor. By W. F. Willoughby and Clare de Graffenried.
   Pp. 150. .75
3-4. Papers on the Canal Question. By E. J. James and L. M. Haupt. Pp. 85. 1.00
5. **History of the New York Property Tax. By J. C. Schwab. Pp. 108. 1.00
6. **Educational Value of Political Economy. By S. N. Patten. Pp. 36. .75

Volume VI, 1891

1-2. Fourth Annual Meeting: Reports, Papers, Discussions. 1.00
3. **Government Forestry. Papers by Pinchot, Bowers, and Fernow. Pp. 102. .75
4-5. **Municipal Ownership of Gas in the U.S. By E. W. Bemis. Pp. 185. 1.00
6. State Railroad Commissions. By F. C. Clark. Pp. 110. .75

Volume VII, 1892

1. **The Silver Situation in the United States. By F. W. Taussig. Pp. 118. .75
2-3. **Shifting and Incidence of Taxation. By E. R. A. Seligman. Pp. 424. (Re-
   vised.) 2.00
4-5. **Sinking Funds. By Edward A. Ross. Pp. 106. 1.00
6. **The Reciprocity Treaty with Canada of 1854. By F. E. Haynes. Pp. 70. .75

Volume VIII, 1893

1. Fifth Annual Meeting: Report of the Proceeding. Pp. 130. .75
2-3. Housing of the Poor in American Cities. By M. T. Reynolds. Pp. 132. 1.00
4-5. Public Assistance of the Poor in France. By E. G. Balch. Pp. 180. 1.00
6. **First Stages of the Tariff Policy of the U.S. By William Hill. Pp. 162. 1.00

Volume IX, 1894

Sixth Annual Meeting: Handbook and Report. Pp. 73. .50
1-2. **Progressive Taxation in Theory and Practice. By Edwin R. A. Seligman.
   Pp. 222. (See 1908, No. 4.) 1.00
3. **The Theory of Transportation. By C. H. Cooley. Pp. 148. .75
4. **Sir William Petty. By Wilson Lloyd Bevan. Pp. 102. .75
5-6. **Papers on Labor Problems. By J. B. Clark, C. D. Wright, D. R. Dewey,
   A. T. Hadley, and J. G. Brooks. Pp. 94. .75

Volume X, 1895

Seventh Annual Meeting: Handbook and Report. Pp. 183. .50
1-3. **The Canadian Banking System, 1817-1890. By R. M. Breckenridge. Pp. 478. 1.50
4. **Poor Laws of Massachusetts and New York. By John Cummings. Pp. 136. .75
5-6. **Letters of Ricardo to McCulloch, 1816-1823. Edited by J. H. Hollander.
   Pp. 204. 1.75

Volume XI, 1896

1-3. **Race Traits and Tendencies of the American Negro. By F. L. Hoffman.
   Pp. 330. 1.25
4. **Appreciation and Interest. By Irving Fisher. Pp. 110. .75
   General Index to Volumes I-XI (1886-1896). .25

ECONOMIC STUDIES

($2.50 per volume)

Volume I, 1896

Eighth Annual Meeting: Handbook and Report. Pp. 78. .50
1. **The Theory of Economic Progress, by J. B. Clark; The Relation of Changes
   in the Volume of the Currency to Prosperity, by F. A. Walker. Pp. 46. .50
2. The Adjustment of Wages to Efficiency. Three Papers: Gain Sharing, by H. R.
   Towne; The Premium Plan, by F. A. Halsey; A Piece-Rate System, by
   F. W. Taylor. Pp. 83. .50
3. **The Populist Movement. By Frank L. McVey. Pp. 81. .50
4. The Present Monetary Situation. By W. Lexis; translated by John Cummings.
   Pp. 72. .50
5-6. **The Street Railway Problem in Cleveland. By W. R. Hopkins. Pp. 94. .75

Volume II, 1897

Ninth Annual Meeting: Handbook and Report. Pp. 162.	$ .50
1.	**Economics and Jurisprudence. By Henry C. Adams. Pp. 48.	.50
2.	The Saloon Question in Chicago. By John E. George. Pp. 62.	.50
3.	The General Property Tax in California. By C. C. Plehn. Pp. 88.	.50
4.	Area and Population of the United States at the Eleventh Census. By W. F. Willcox. Pp. 60.	.50
5.	**A Discussion Concerning the Currencies of the British Plantations in America, etc. By William Douglass. Edited by C. J. Bullock. Pp. 228.	.50
6.	Density and Distribution of Population in the United States at the Eleventh Census. By W. F. Willcox. Pp. 79.	.50

Volume III, 1898

Tenth Annual Meeting: Handbook and Report. Pp. 136.	.50
1.	**Government by Injunction. By William H. Dunbar. Pp. 44.	.50
2.	**Economic Aspects of Railroad Receiverships. By H. H. Swain. Pp. 118.	.50
3.	The Ohio Tax Inquisitor Law. By T. N. Carver. Pp. 50.	.50
4.	The American Federation of Labor. By Morton A. Aldrich. Pp. 54.	.50
5.	Housing of the Working People in Yonkers. By E. L. Bogart. Pp. 82.	.50
6.	The State Purchase of Railways in Switzerland. By Horace Michelie; translated by John Cummings. Pp. 72.	.50

Volume IV, 1899

Eleventh Annual Meeting: Handbook and Report. Pp. 126.	.50
1.	I., Economics and Politics. By A. T. Hadley. II. Report on Currency Reform. III. Report on the Twelfth Census. Pp. 70.	.50
2.	**Personal Competition. By Charles H. Cooley. Pp. 104.	.50
3.	Economics as a School Study. By F. R. Clow. Pp. 72.	.50
4-5.	The English Income Tax. By J. A. Hill. Pp. 162.	1.00
6.	**Effects of Recent Changes in Monetary Standards upon the Distribution of Wealth. By F. S. Kinder. Pp. 91.	.50

NEW SERIES

1.	**The Cotton Industry. By M. B. Hammond. Pp. 382.	1.50
2.	**Scope and Method of the Twelfth Census. Critical discussion by over twenty statistical experts. Pp. 625.	2.00

THIRD SERIES

Note—During 1896-1899 the Association issued its publications in two series, viz., the bimonthly Economic Studies, *and the "New Series" of larger monographs printed at irregular intervals. In 1900 it reverted to the policy of issuing its monographs, now called the "Third Series" of the publications, at quarterly intervals.*
· Price per volume, $4.00.

Volume I, 1900

1.	Twelfth Annual Meeting: Papers on Economic Theory and Political Morality; Trusts; Railroad Problems; Public Finance; Consumers' League; Twelfth Census. Pp. 186.	1.00
2.	**The End of Villeinage in England. By T. W. Page. Pp. 99.	1.00
3.	**Essays in Colonial Finance. By Members of the Association. Pp. 303.	1.50
4.	**Currency and Banking in the Province of Massachusetts Bay. By A. McF. Davis. Part I: Currency. Pp. 464 + 91 photogravure plates.	1.75

Volume II, 1901

1.	**Thirteenth Annual Meeting; Papers on Competition; Commercial Education; Economic Theory; Taxation of Quasi-public Corporations; Municipal Accounts. Pp. 300.	1.25
2.	**Currency and Banking. By A. McF. Davis. Part II: Banking. Pp. 341 + 18 photogravure plates.	1.75
3.	**Theory of Value before Adam Smith. By Hannah R. Sewall. Pp. 132.	1.00
4.	Administration of City Finances in the U.S. By F. R. Clow. Pp. 144.	1.00

Volume III, 1902

1. **Fourteenth Annual Meeting: Papers on International Trade; Industrial Policy; Public Finance; Protective Tariff; Negro Problem; Arbitration of Labor Disputes in Coal Mining Industry; Porto Rican Finance; Economic History. Pp. 400. $1.50
2. **The Negro in Africa and America. By Joseph A. Tillinghast. Pp. 240. 1.25
3. Taxation in New Hampshire. By M. H. Robinson. Pp. 232. 1.25
4. **Rent in Modern Economic Theory. By Alvin S. Johnson. Pp. 136. .75

Volume IV, 1903

1. Fifteenth Annual Meeting: Papers and Discussions on Economic and Social Progress; Trade Unions and the Open Shop; Railway Regulation; Theory of Wages; Theory of Rent; Oriental Currency Problem. Pp. 298. 1.25
2. Ethnic Factors in the Population of Boston. By F. A. Bushee. Pp. 171. 1.00
3. History of Contract Labor in the Hawaiian Islands. By Catharine Coman. Pp. 74. .75
4. **The Income Tax in the Commonwealth of the United States. By Delos O. Kinsman. Pp. 134. 1.00

Volume V, 1904

**Sixteenth Annual Meeting. Papers and Proceedings published in two parts.
1. **Part I—Papers and Discussions on Southern Agricultural and Industrial Problems; Social Aspects of Economic Law; Relations between Rent and Interest. Pp. 240. 1.00
2. **Part II—Papers and Discussions on The Management of the Surplus Reserve; Theory of Loan Credit in Relation to Corporation Economics; State Taxation of Interstate Commerce; Trusts; Theory of Social Causation. Pp. 203. 1.00
3. Monopolistic Combinations in the German Coal Industry. By Francis Walker. Pp. 340. 1.25
4. **The Influence of Farm Machinery on Production and Labor. By Hadley Winfield Quaintance. Pp. 110. .75

Volume VI, 1905

Seventeenth Annual Meeting. Papers and Proceedings published in two parts.
1. Part I—Papers and Discussions on the Doctrine of Free Trade; Theory of Prices; Theory of Money; Open Shop or Closed Shop. Pp. 226. 1.00
2. Part II—Papers and Discussions on Government Interference with Industrial Combinations; Regulation of Railway Rates; Taxation of Railways; Preferential Tariffs and Reciprocity; Inclosure Movement; Economic History of the United States. Pp. 270. 1.00
3. **The History and Theory of Shipping Subsidies. By R. Meeker. Pp. 230. 1.00
4. Factory Legislation in the State of New York. By F. R. Fairchild. Pp. 218. 1.00

Volume VII, 1906

1. Eighteenth Annual Meeting: Papers and Discussions on The Love of Wealth and the Public Service; Theory of Distribution; Governmental Regulation of Railway Rates; Municipal Ownership; Labor Disputes; The Economic Future of the Negro. Pp. 325. 1.00
2. **Railroad Rate Control. By H. S. Smalley. Pp. 147. 1.00
3. On Collective Phenomena and the Scientific Value of Statistical Data. By E. G. F. Gryzanovski. Pp. 48. .75
   Handbook of the Association, 1906. Pp. 48. .25
4. The Taxation of the Gross Receipts of Railways in Wisconsin. By G. E. Snider. Pp. 138. 1.00

Volume VIII, 1907

1. Nineteenth Annual Meeting: Papers and Discussions on Modern Standards of Business Honor; Wages as Determined by Arbitration; Commercial Education; Money and Banking; Western Civilization and Birth Rate; Economic History; Government Regulation of Insurance; Trusts and Tariff; Child Labor. Pp. 268. 1.00
2. **Historical Sketch of the Finances and Financial Policy of Massachusetts from 1780 to 1905. By C. J. Bullock. Pp. 144. 1.00
   Handbook of the Association, 1907. Pp. 50. .25

3. **The Labor Legislation of Connecticut. By Alba M. Edwards. Pp. 322. $1.00
4. The Growth of Large Fortunes. By G. P. Watkins. Pp. 170. 1.00

Volume IX, 1908

Handbook of the Association, 1908. Pp. 49. .25
1. Twentieth Annual Meeting: Papers and Discussions on Principles of Government Control of Business; Are Savings Income; Agricultural Economics; Money and Banking; Agreements in Political Economy; Labor Legislation; Relation of the Federal Treasury to the Money Market; Public Service Commissions. Pp. 311. 1.25
2. Chicago Traction. By R. E. Heilman. Pp. 131. 1.00
3. Factory Legislation of Rhode Island. By J. K. Towles. Pp. 119. 1.00
4. **Progressive Taxation in Theory and Practice. Revised Edition. By E. R. A. Seligman. Pp. 334. 1.25

Volume X, 1909

1. **Twenty-first Annual Meeting: Papers and Discussions on The Making of Economic Literature; Collective Bargaining; Round Table on Accounting; Labor Legislation; Employers' Liability; Canadian Industrial-Disputes Act; Modern Industry and Family Life; Agricultural Economics; Transportation; Revision of the Tariff; A Central Bank; The National Monetary Commission; Capitalization of Public Service Corporations in Massachusetts. Pp. 432. 1.50
2. Handbook of the Association, 1909. Pp. 59. .25
3. **The Printers. By George E. Barnett. Pp. 379. 1.50
4. Life Insurance Reform in New York. By W. H. Price. Pp. 95. .75

Volume XI, 1910

1. **Twenty-second Annual Meeting: Papers and Disscussions on History of the Association; Observation in Economics; Economic Dynamics; Theory of Wages; Country Life; Valuation of Public Service Corporations; Trusts; Taxation. Pp. 386. 1.50
2. **Handbook of the Association, 1910. Pp. 79. .25
3. The Child Labor Policy of New Jersey. By A. S. Field. Pp. 229. 1.25
4. The American Silk Industry and the Tariff. By F. R. Mason. Pp. 178. 1.00

THE ECONOMIC BULLETIN

Published quarterly in 1908, 1909, and 1910, and containing personal notes, news of the economic world, announcements of special lines of investigation, and a classified and annotated bibliography of the current books and articles on economic subjects.

Volume   I, 1908, Four numbers. $2.00; each .60
Volume II, 1909. Four numbers. $2.00; each .60
Volume III, 1910. Four numbers. $2.00; each .60

FOURTH SERIES

Volume I, 1911

The American Economic Review, March, June, September, and December
Twenty-third Annual Meeting:
> Papers and Discussions on the Significance of a Comprehensive System of Education; Money and Prices; The Ricardo Centenary; Accounting; Canals and Railways; Population and Immigration; Labor Legislation; Taxation; A Definition of Socialism; Competition in the Fur Trade. Pp. 388.

Handbook of the Association, 1911.

Volume II, 1912

The American Economic Review, March, June, September, and December
Supplement.—Twenty-fourth Annual Meeting:
> Papers and Proceedings. The Economic Utilization of History; Tariff Legislation; The Federal Budget; Rural Conditions; Selection of Population by Migration; The Price Concept; An International Commission on the Cost of Living; Industrial Efficiency. Pp. 146.

Volume III, 1913
The American Economic Review, March June, September, and December
Supplement.—Twenty-fifth Annual Meeting:
     Papers and Proceedings. Population or Prosperity; The Rising Cost of Liv-
     ing—Standardizing the Dollar; Banking Reform; Theories of Distribution;
     Farm Management; Governmental Price Regulation. Pp. 155.
**Supplement.—Handbook of the Association, 1913.

Volume IV, 1914
The American Economic Review, March, June, September, and December
Supplement.—Twenty-sixth Annual Meeting:
     Papers and Proceedings. The Increasing Governmental Control of Economic
     Life; The Control of Public Utilities; Railroad Rate Making; Syndicalism;
     Trust Decisions and Business. Pp. 211.

Volume V, 1915
The American Economic Review, March, June, September, and December
Supplement.—Twenty-seventh Annual Meeting:
     Papers and Proceedings. Economics and the Law; Regulation of the Stock
     Exchanges; Market Distribution; Statistical Work of the United States
     Government; Relation of Education to Industrial Efficiency; The Effect of
     Income and Inheritance Taxes on the Distribution of Wealth; Public Regu-
     lation of Wages. Pp. 323.

Volume VI, 1916
The American Economic Review, March, June, September, and December
Supplement.—Twenty-eighth Annual Meeting:
     Papers and Proceedings. The Apportionment of Representatives; Effect of
     the War on Foreign Trade; Budget Making and the Increased Cost of
     Government; Economic Costs of War; Economic Theorizing and Scientific
     Progress; The Role of Money in Economic Theory; Price Maintenance;
     Investment of Foreign Capital. Pp. 248.
**Supplement.—Handbook of the Association, 1916.

Volume VII, 1917
The American Economic Review, March, June, September, and December
Supplement.—Twenty-ninth Annual Meeting:
     Papers and Proceedings. The National Point of View in Economics; Landed
     Property; Two Dimensions of Economic Productivity; Some Social Surveys
     in Iowa; The Land Problem and Rural Welfare; The Federal Farm Loan
     Act; Statistics of the Concentration of Wealth; Gold Supply at the Close
     of the War; Loans and Taxes in War Finance; Problems of Population
     after the War; Some Phases of the Minimum Wage. Pp. 275.
Supplement.—Index to the Publications, 1886-1910.

Volume VIII, 1918
The American Economic Review, March, June, September, and December
Supplement.—Thirtieth Annual Meeting:
     Papers and Proceedings. Economic Reconstruction; Federal Taxes upon In-
     come and Excess Profits; Land Utilization and Colonization; Federal Val-
     uation of Railroads; Co-ordination of Employment Bureaus; Control of
     the Acquisition of Wealth; Motives in Economic Life; Price-Fixing Prob-
     lems of Governmental Efficiency; Economic Alliances and Tariff Adjust-
     ments. Pp. 317.

Volume IX, 1919
The American Economic Review, March, June, September, and December
Supplement.—Thirty-first Annual Meeting:
     Papers and Proceedings. Economists in Public Service; Interest on Invest-
     ment a Manufacturing Cost Factor; Control of Marketing Methods and
     Costs; War and the Supply of Capital; War and the Rate of Interest;
     Index of the Cost of Living; Securing the Initiative of the Workman; A
     Legal Dismissal Wage; After-War Gold Policies; Foreign Exchange; Sta-
     bilizing the Dollar; Tenancy of Landed Property; Price-Fixing; Economic

Theory in an Era of Readjustment; Psychology and Economics; The Open Door and Colonial Policy; Reports of Committees on Foreign Trade and Purchasing Power of Money. Pp. 368.

Supplement No. 2—Report of the Committee on War Finance. Pp. 142.

**Supplement No. 3—Handbook of the Association, 1919.

Volume X, 1920

The American Economic Review, March, June, September, and December

Supplement.—Thirty-second Annual Meeting:

Papers and Proceedings. Excess Profits Taxation; Germany's Reparation Payments; International Supervision over Foreign Investments; Results of a Balanced Industrial System; Employee's Representation in Management of Industry; Prices and Reconstruction; Banking Policy and Prices; Large-Scale Marketing; Reports of Committees on Foreign Trade, Co-ordination in Taxation, Census Advisory Committee. Pp. 278.

Supplement No. 2—Taxation of Excess Profits in Great Britain. Pp. 244.

Volume XI, 1921

The American Economic Review, March, June, September, and December

Supplement.—Thirty-third Annual Meeting:

Papers and Proceedings. The Railroad Situation; Our Foreign Trade Balance; Problems of the Bituminous Coal Industry; Traditional Economic Theory; Non-Euclidean Economics; Federal Taxation of Profits and Income; Teaching of Elementary Economics. Pp. 194.

Supplement No. 2.—Karelsen Prize Essays, on What Can a Man Afford? Pp. 118.

Volume XII, 1922

The American Economic Review, March, June, September, and December

Supplement.—Thirty-fourth Annual Meeting:

Papers and Proceedings. Business Cycles; American Trade Unionism; The Railroads; Workmen's Compensation; Federation in Central America; Teaching of Elementary Economics; The Chain Store Grocer; Economics and Ethics. Pp. 194.

**Supplement No. 2—Handbook of the Association, 1922.

Volume XIII, 1923

The American Economic Review, March,   June, September, and December

Supplement.—Thirty-fifth Annual Meeting:

Papers and Proceedings. Trend of Prices; Trend of Rate of Interest and Investment; Outlook for Wages and Employment; Overhead Costs; Commercial Rent and Profits; Labor Turnover; Factors in Wage Determinations; Income of Farmers; Large-Scale Production and Merchandising; Marketing Farm Products; Bureau of Business Research. Pp. 293.

Volume XIV, 1924

The American Economic Review, March, June, September, and December

Supplement.—Thirty-sixth Annual Meeting:

Papers and Proceedings. International Trade and Commercial Policy; Railroad Consolidation; Economic Theory; Transportation; American Foreign Trade; Marketing. Pp. 192.

**Supplement No. 2—Handbook of the Association, 1924.

Volume XV, 1925

The American Economic Review, March, June, September, and December

Supplement.—Thirty-seventh Annual Meeting:

Papers and Proceedings. The Economics of Advertising; Problems of Economic Theory; Transportation; Marketing; Giant Power; The Teaching of Business and Economics; Business Administration; Monetary Stabilization; Foreign Service Training; Highway Economics; Psychological Problems of Industry. Pp. 165.

Supplement No. 2.—Babson Prize Essay, on Forecasting the Price of Hogs. Pp. 22.

Volume XVI, 1926

The American Economic Review, March, June, September, and December

Supplement.—Thirty-eighth Annual Meeting:
Papers and Proceedings. Movement of Real Wages; Teaching of Economics; Consuming Power of Labor and Business Fluctuations; Economic Problems Involved in the Payment of International Debts; Economics and Geography; Agriculture in Our National Policy; Tariff Making; Trade Associations; Theory of Wages; Reducing the Costs of Marketing; Topics in Economic History; Railway Problems; Land Economics; Federal Reserve Policies. Pp. 353.

**Supplement No. 2.—Handbook of the Association, 1926.

Volume XVII, 1927

The American Economic Review, March, June, September, and December
Supplement.—Thirty-ninth Annual Meeting:
Papers and Proceedings. Economics of Prohibition; Economic History; Use of the Quantitative Method in the Study of Economic Theory; Present-Day Corporation Problems; American Practices Analogous to Foreign Controls over Raw Materials; Marketing; Interest Theory and Price Movements; Problem of Effective Public Utility Regulation; Immigration Restriction—Economic Results and Prospects; Family Budgets; Motor Transportation in the United States. Pp. 218.

Supplement No. 2.—Report of the Dinner in Honor of Professor John Bates Clark. Pp. 18.

Volume XVIII, 1928

The American Economic Review, March, June, September, and December
Supplement.—Fortieth Annual Meeting:
Papers and Proceedings. Land Economics; Marketing; Present Status and Future Prospects of Quantitative Economics; Post-War Fluctuations of Commodity Prices; Relationship between Departments of Economics and Collegiate Schools of Business; Economic History; Simplification of the Federal Income Tax; Economics Significance of the Increased Efficiency of American Industry; An Approach to the Law of Production and Its Relation to the Welfare of the Wage-Earner; Meaning of Valuation; Railroad Valuation with Special Reference to the O'Fallon Decision; Interest Rates As Factors in the Business Cycle; Should the Debt Settlements Be Revised; An Examination of the Reasons for Revision of the Debt Settlements. Pp. 305.

Supplement No. 2.—Handbook of the Association, 1928.

Volume XIX, 1929

The American Economic Review, March, June, September, and December
Supplement.—Forty-first Annual Meeting:
Papers and Proceedings. Market Shifts, Price Movements, and Employment; Some Observations on Unemployment Insurance; Marketing; Land Economics; Law and Economics; Price Stabilization; London and the Trade Cycle; Federal Reserve Policy and Brokers' Loans; Central Planning of Production in Soviet Russia; International Differences in the Labor Movement; Tariff Making in the United States; Economic History; Locality Distribution of Industries; Regulation of Electric Light and Power Utilities; An Inductive Study of Publicly Owned and Operated vs. Privately Owned but Regulated Public Utilities; Regulation of the Common Carrier; Commercial Motor Vehicle and the Public. Pp. 284.

Volume XX, 1930

The American Economic Review, March, June, September, and December
Supplement.—Forty-second Annual Meeting:
Papers and Proceedings, Economic History; Public Works Plan and Unemployment; Theory of Economic Dynamics as Related to Industrial Instability; Chief Economic Problems of Mexico; Reparations Settlement and the International Flow of Capital; Federal Reserve Board—Its Problems and Policy; Economic and Social Consequences of Mechanization in Agriculture and Industry. Pp. 214.

Volume XXI, 1931

The American Economic Review, March, June, September, and December
Supplement.—Forty-third Annual Meeting:
Papers and Proceedings. Decline of Laissez Faire; Small Loan Business; So-
cial and Economic Aspects of Chain Stores; Russian Economic Situation;
Trustification and Economic Theory; Persistence of the Merger Movement;
Program of the Federal Farm Board; Social Implications of Restriction of
Agricultural Output; Land Economics and Real Estate; Institutionalism—
What It Is and What It Hopes to Become; An Approach to World Eco-
nomics; International Industrial Relations—Migration of Enterprise and
Policies Affecting It; World-Wide Depression of 1930; Present Depres-
sion—A Tentative Diagnosis; Power and Propaganda; Failure of Electric
Light and Power Regulation and Some Proposed Remedies. Pp. 302.
**Supplement No. 2.—Handbook of the Association, 1931.

Volume XXII, 1932

The American Economic Review, March, June, September, and December
Supplement.—Forty-fourth Annual Meeting:
Papers and Proceedings. Private Enterprise in Economic History; Shorter
Working Time and Unemployment; Quantitative Economics; Theory of
Technological Progress and the Dislocation of Employment; Measurement
of Productivity Changes and the Displacement of Labor; Stabilization of
Business and Employment; Principle of Planning and the Institution of
Laissez Faire; Institutional Economics; Elasticity of Demand as a Useful
Marketing Concept; Investments of Life Insurance Companies; Real Estate
in the Business Cycle; Investments and National Policy of the United
States in Latin America; Recent Changes in the Character of Bank Lia-
bilities and the Problem of Bank Reserves; Bank Failures in the United
States; Transportation by Rail and Otherwise; Our Changing Transporta-
tion System. Pp. 306.

Volume XXIII, 1933

The American Economic Review, March, June, September, and December
Supplement.—Forty-fifth Annual Meeting:
Papers and Proceedings. Rise of Monopoly in the United States; Record of
Insurance in the Depression; Some Theoretical Aspects of Unemployment
Reserves; The Economics of Unemployment Relief; American Economic
Thought; Formation of Capital; Measurement and Relation to Economic
Instability; Size of Business Unit as a Factor in Efficiency of Marketing;
Reserve Bank Policy and Economic Planning; Federal Reserve Policy in
World Monetary Chaos; Tariff Reform: The Case for Bargaining; Specula-
tion in Suburban Lands; Real Estate Speculation and the Depression. Pp.
206.
**Supplement No. 2.—Handbook of the Association, 1933.

Volume XXIV, 1934

The American Economic Review, March, June, September, and December
Supplement.—Forty-sixth Annual Meeting:
Papers and Proceedings. The History of Recovery; Public Utilities in the
Depression; Imperfect Competition; Fundamentals of a National Transpor-
tation Policy; Correlation of Rail and Highway Transportation; Marketing
under Recovery Legislation; Economics of the Recovery Act; Measurement
of Unemployment; Controlled Inflation; Banking Act of 1933—An Ap-
praisal; Some Statistics on the Gold Situation; The Problem of Tax De-
linquency; The Problem of Expanding Governmental Activities; The Eco-
nomics of Public Works. Pp. 224.

Volume XXV, 1935

The American Economic Review, March, June, September, and December
Supplement.—Forty-seventh Annual Meeting:
Papers and Proceedings. NRA Examined; Rate-making Problems of TVA;
New Deal and the Teaching of Economics; Paths of Economic Change;
Business Enterprise and the Organization of Production; Changes in the

Character, Structure, and Conditions of Production; International Aspects of Problems of Production and Trade; International Movements of Capital; Our Commercial Banking System; Aspects of Co-ordination and Finance; Some Lessons Drawn from European Experience; Nationalism; Security Regulation and Speculation; Monetary Stabilization from an International Point of View; Monetary Stabiliation from a National Point of View; Decentralization of Population and Industry; Co-ordination of State and Local Finance; Relief Aspects of the New Deal; Unified Program for the Unemployed. Pp. 240.

Volume XXVI, 1936

The American Economic Review, March, June, September, and December.
Supplement.—Forty-eighth Annual Meeting:
Papers and Proceedings. Some Distinguishing Characteristics of the Current Recovery; Price Theories and Market Realities; Notes on Inflexible Prices; Effect of the Depression upon Earnings and Prices of Regulated and Nonregulated Industries; Size of Plants in Its Relation to Price Control and Price Flexibility; Requisites of Free Competition; Monopolistic Competition and Public Policy; Banking Act of 1935; Recent Legislation and the Banking Situation; Economic Aspects of an Integrated Social Security Program; Capital Formation; Trade Agreements Program and American Agriculture; Founding and Early History of the American Economic Association; Developments in Economic Theory; Federal Revenue Act of 1935; Relations between Federal, State, and Local Finances; Equalization of Local Government Resources; Adjustment to Instability; Transportation Problems; Fifty Years' Developments in Ideas of Human Nature and a Motivation; Institutional Economics; Place of Marginal Economics in a Collectivist System; Problem of Prices and Valuation in the Soviet System; Effects of New Deal Legislation on Industrial Relations; Report of the Fiftieth Anniversary Dinner. Pp. 350.
**Supplement No. 2.—Handbook of the Association, 1936.

Volume XXVII, 1937

The American Economic Review, March, June, September, and December
Supplement.—Forty-ninth Annual Meeting:
Papers and Proceedings. Economic Interdependence, Present and Future; Quantitative and Qualitative Changes in International Trade During the Depression; Current Tendencies in Commercial Policy; Trade Problem of the Pacific; Analysis of the Nature of American Public Debts; Limits to Possible Debt Burdens, Federal, State, and Local; Debt Retirement and the Budget; United States Debt—Distribution among Holders and Present Status; Federal-State Unemployment Compensation Provisions of the Social Security Act; Unemployment Relief and Insurance; Economic Problems Arising from Social Security Taxes and Reserves; The Situation of Gold Today in Relation to World Currencies; Mechanisms and Objectives for the Control of Exchange; The Adequacy of Existing Currency Mechanisms Under Varying Circumstances; Present Situation of Inadequate Housing; Financing of Housing; Some Economic Implications of Modern Housing; Managed Currency; A Critique of Federal Personnel Policies as Applied to Professional Social Science Positions; New Opportunities for Economists and Statisticians in Federal Employment; Government Employment as a Professional Career in Economics; Indicia of Recovery; Housing and Housing Research; Distribution of Purchasing Power and Business Fluctuations; Forecast of Power Development; The Possibility of a Scientific Electrical Rate System; Co-ordination of Public and Private Power Interests in European Countries; Recent Developments in the Theory of Speculation; Control of Speculation under the Securities Exchange Act; Unorganized Speculation: the Possibility of Control. Pp. 333.

Volume XXVIII, 1938

The American Economic Review, March, June, September, and December
Supplement.—Fiftieth Annual Meeting:
Papers and Proceedings. The Significance of Marxian Economics for Present-day Economic Theory; The Significance of Marxian Economics for

Current Trends of Governmental Policy; The Rate of Interest; Security
Markets and the Investment Process; Relation of Price Policy to Fluctua-
tions of Investment; General Interest Theory; Rate of Interest; Security
Regulation; Corporate Price Policies; Fiscal Policies; Rate of Consump-
tion; Wage Rates; Social Security Program; Rate of Consumption; Dur-
able Consumers Goods; Wage Policies. Pp. 192.
Supplement No. 2.—Handbook of the Association, 1938.

Volume XXIX, 1939
The American Economic Review, March, June, September, and December
Supplement.—Fifty-first Annual Meeting:

Papers and Proceedings. Problem of Industrial Growth in a Mature Econ-
omy; Effects of Current and Prospective Technological Developments upon
Capital Formation; Public Investment in the United States; Expansion and
Contraction in the American Economy; Effect of Industrial and Techno-
logical Developments upon Demand for Capital; Role of Public Investment
and Consumer Capital Formation; Income and Capital Formation; Price
and Production Policies of Large-Scale Enterprise; Changing Distribution
Channels; Financial Control of Large-Scale Enterprise; Pure Theory of
Production; Changing Character of American Industrial Relations; Wages
and Hours in Relation to Innovations and Capital Formation; Effect of
Wage Increase upon Employment; Relation of Wage Policies and Price
Policies; An Appraisal of Factors Which Stopped Short the Recovery De-
velopment in the United States; Fiscal Policy in the Business Cycle; An
Appraisal of the Workability of Compensatory Devices; Divergencies in the
Development of Recovery in Various Countries; Factors Making for
Change in Character of Business Cycle; Industrial Relations. Pp. 280.

Volume XXX, 1940
The American Economic Review, March, June, September, and December
Supplement.—Fifty-second Annual Meeting:

Papers and Proceedings. Objectives of Monetary Policy; Economic Issues
in Social Security Policy; Bank Deposits and the Business Cycle; Problems
in the Teaching of Economics; Price Control Under "Fair Trade" Legisla-
tion; Problems of American Commercial Policy; Transportation Problem;
Preserving Competition Versus Regulating Monopoly; Theory of Interna-
tional Trade; Collective Bargaining and Job Security; Banking Reform
Through Supervisory Standards; Incidence of Taxation; Economic Plan-
ning; Growth of Rigidity in Business; Economics of War; Population
Problems; Cost Functions and Their Relation to Imperfect Competition.
Pp. 436.
Supplement No. 2.—Handbook of the Association, 1940.
No. 5 (February, 1941)
Fifty-third Annual Meeting (December, 1940):

Papers and Proceedings. Gold and the Monetary System; Economic Re-
search; Federal Budget; Economic Consequences of Deficit Financing;
Teaching of Economics; Agricultural Situation; A Review of Fundamental
Factors, an Evaluation of Public Measures, and an Appraisal of Prospects;
Status and Role of Private Investment in the American Economy, 1940;
Unemployment in the United States, 1930-50; Economic Consequences of
War Since 1790; Some Economic Problems of War, Defense, and Postwar
Reconstruction; United States in the World Economy, 1940; International
Economic Relations and Problems of Commercial Policy; Price Policy and
Price Behavior. Pp. 458.

Volume XXXI, 1941
The American Economic Review, March, June, September, and December

Volume XXXII, 1942
The American Economic Review, March, June, September, and December
Supplement.—Fifty-fourth Annual Meeting:

Papers and Proceedings. Economic Adjustments After Wars; Problems of
Taxation; Determinants of Investment Decisions; Problems of Interna-

tional Economic Policy for the United States; History of American Corporations; Problems of Labor Market Research; Co-ordination of Federal, State, and Local Fiscal Policy; Technical Aspects of Applying a Dismissal Wage to Defense Workers; Problems of International Economic Policy; Impact of National Defense and the War upon Public Utilities; Future of Interest Rates; Effect of Managerial Policy upon the Structure of American Business; Economic Effects of Wars; Economic Aspects of Reorganization Under the Chandler Act; Economics of Industrial Research; Objectives in Applied Land Economics Curricula; Changing Position of the Banking System and Its Implications for Monetary Policy; Determination of Wages; Economic Problems of American Cities; Cost and Demand Functions of the Individual Firm; Problems of Price Control; Effects of the War and Defense Program upon Economic Conditions and Institutions; Trade Unions and the Law. Pp. 534.

Supplement No. 2.—Papers Relating to the Temporary National Economic Committee. Pp. 135.

Supplement No. 3.—Directory. Pp. 198.

Volume XXXIII, 1943

The American Economic Review, March, June, September, and December
Supplement.—Fifty-fifth Annual Meeting:

Papers and Proceedings. Economic Claims of Government and of Private Enterprise; Our Industrial Plant When Peace Comes; Financial and Government Contract Adjustments of Industry at the End of the War; Problems of Public Policy Raised by Collective Bargaining; Our Labor Force When Peace Comes; Price Control and Rationing; Case Studies in Price Control; Restoration of International Trade; Future of International Investment; International Financial Relations After the War; Economic Regionalism and Multilateral Trade; Bases of International Economic Relations; International Commodity Agreements. Pp. 508 + 15.

Volume XXXIV, 1944

The American Economic Review, March, June, September, and December
Supplement.—Fifty-sixth Annual Meeting:

Papers and Proceedings. Political Science, Political Economy, and Values; Educational Function of Economists and Political Scientists; Public Administration of Transportation under War Conditions; How Achieve Full and Stable Employment; Incentive Problems in Regulated Capitalism; Postwar Labor Problems; Social Security; Postwar Legal and Economic Position of American Women; Postwar Domestic Monetary Problems; Economic Organization of Welfare; International Trade; Regional Problems; International Monetary Problems. Pp. 440 + 16.

Supplement No. 2.—Implemental Aspects of Public Finance. Pp. 138.

Volume XXXV, 1945

The American Economic Review, March, June, September, and December
Supplement (May).—Fifty-seventh Annual Meeting:

Papers and Proceedings. Consumption Economics; Expanding Civilian Production and Employment After the War; Natural Resources and International Policy; Interdepartmental Courses in the Social Sciences; Price Control and Rationing in the War-Peace Transition; Organized Labor and the Public Interest; Aviation in the Postwar World; International Monetary and Credit Arrangements; Agricultural Price Supports and Their Consequences; Political Economy of International Cartels; Fiscal Problems of Transition and Peace; Problems of Regionalism in the United States; Food and Agriculture—Outlook and Policy; Function of Government in the Postwar American Economy. Pp. 520 + 16.

Volume XXXVI, 1946

The American Economic Review, March, June, September, and December
Supplement (May).—Fifty-eighth Annual Meeting:

Papers and Proceedings. Problem of "Full Employment"; American Economy in the Interwar Period; Postwar Labor Relations; Monetary Policy;

Changing Structure of the American Economy; Economic Problems of Foreign Areas; Publication of an Annual Review of Economics; New Frontiers in Economic Thought; Postwar Shipping Policy; Monopoly and Competition; Postwar Tax Policy; Postwar Railroad Problems; International Investment; Recent Developments in Public Utility Regulation; International Cartels; Economic Research; Methods of Focusing Economic Opinion on Questions of Public Policy (e.g., Monetary, Agricultural Price Supports); Undergraduate Teaching of Economics. Pp. 960.

**Supplement No. 2.—Handbook. Pp. 143.

Volume XXXVII, 1947

The American Economic Review, March, June, September, and December
Supplement (May).—Fifty-ninth Annual Meeting:
Papers and Proceedings. Employment Act of 1946 and a System of National Bookkeeping; Social and Economic Significance of Atomic Energy; Public Debt: History, Effects on Institutions and Income, and Monetary Aspects; Economic Forecasts; Role of Social Security in a Stable Prosperity; Economic Outlook; Economy of the U.S.S.R.; Domestic versus International Economic Equilibrium; Prices: Wartime Heritage and Some Present Problems; Banking Problems; Productivity in the American Economy; International Trade Organization; Vital Problems in Labor Economics; Transportation and Public Utilities Problems; Housing Problems; Economic Research; Changing Character of Money. Pp. 781.

Volume XXXVIII, 1948

The American Economic Review, March, June, September, and December
Supplement (May).—Sixtieth Annual Meeting:
Papers and Proceedings. Economic Theory of Imperfect Competition, Oligopoly, and Monopoly; Role of Monopoly in the Colonial Trade and Expansion of Europe; Progress of Concentration in Industry; Does Large-Scale Enterprise Result in Lower Costs; Sherman Act and the Enforcement of Competition; Patent Policy; A Consideration of the Economic and Monetary Theories of J. M. Keynes; Keynesian Economics: The Propensity to Consume and the Multiplier, and Savings, Investment, and Wage Rates; Economics Collides with Ethics; An Appraisal of the Taft-Hartley Act; Fiscal Policy in Prosperity and Depression; Problems of Timing and Administering Fiscal Policy in Prosperity and Depression; Transportation and Public Utilities; Futility of Trust-Busting; National Productivity. Pp. 591.

Volume XXXIX, 1949

The American Economic Review, March, June, September, and **December
Supplement (No. 1—January).—Directory. Pp. 343.
Supplement (No. 3—May).—Sixty-first Annual Meeting:
Papers and Proceedings. Commemoration of the Centenary of the Communist Manifesto—The Sociology and Economics of Class Conflict; Current Research in Business Cycles; Interregional Variations in Economic Fluctuations; Economic Research; Economic Consequences of Some Recent Antitrust Decisions; Theory and Measurement of Price Expectations; Input-Output Analysis and Its Use in Peace and War Economies; Liquidity and Uncertainty; Problems of the ITO; Commodity Marketing—Going Where; John Stuart Mill—Centennial Appraisal; Possibilities for a Realistic Theory of Entrepreneurship; Economics of Preparedness for War; Present Issues of the Latin-American Economy. Pp. 537.

Volume XL, 1950

The American Economic Review, March, June, September, and December
Supplement (No. 2—May).—Sixty-second Annual Meeting:
Papers and Proceedings. What Planning and How Much Is Compatible with a Market Economy—Recent European Experience; Capitalism and Monopolistic Competition—I. The Theory of Oligopoly, II. Can the American Economy Be Made More Competitive; Capitalism and Economic

Progress; Stabilizing the Economy—The Employment Act of 1946 in Operation; Problems of an Advanced Defense Economy; Transportation in Capitalist and Socialized Economies; Can Capitalism Dispense with Free Labor Markets; Capitalism and Equality of Income; Tax Structure and Private Enterprise; Economic Power Blocs and American Capitalism; American Capitalism—Where Are We Going; U. S. Foreign Investment in Underdeveloped Areas; Economic Policy in Occupied Germany. Pp. 650.

Supplement.—Supplement to 1948 Directory. Pp. 41.

Supplement No. 2.—The Teaching of Undergraduate Economics. Pp. 226.

Volume XLI, 1951

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-third Annual Meeting:

Papers and Proceedings. Role and Interests of the Consumer; Institutional Economics; Executive Decisions at the Top Level; Economic Theory, Statistics, and Economic Practice; Capital and Interest; Economic Stabilization; Modern Industrialism and Human Values; Factors in Modern Industrial Development; Government Action or Private Enterprise in River Valley Development; Economic Progress—Some Cases, Comparisons, and Contrasts; Point Four—Development of Backward Areas; Current Problems in International Trade; Economy of the Soviet Union; Changes in National Transportation Policy; Public Utilities and National Policy; Housing Problem—Current Situation and Long-run Effects of Government Housing Programs; Economics of Medical Care—The Problem and Alternative Solutions; Economics in General Education. Pp. 816.

Supplement.—Supplement to 1948 Directory. Pp. 17.

Volume XLII, 1952

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-fourth Annual Meeting:

Papers and Proceedings. Economic Theory and Public Policy; Issues in Methodology; Business Cycle Theory; Monetary Theory; Fiscal Theory; International Trade Theory; Value Theory; Recent Developments in United States Monetary Policy; Inflation Control in the United States; American Foreign Aid Programs; International Trade in the Postwar World; Wages, Manpower, and Rearmament; Governmental Policy on Business Practices; Economic Problems of Military Mobilization; Theoretical Analysis of Economic Growth; General Factors in Economic Growth in the United States; Growth in Underdeveloped Countries; Role of War in American Economic Development; Public Utilities, Transportation, and Spatial Organization; Collective Bargaining in the Regulated Industries. Pp. 768.

Supplement.—Supplement to 1948 Directory. Pp. 11.

Volume XLIII, 1953

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-fifth Annual Meeting:

Papers and Proceedings: Monetary Policy; A Stock-taking of Bretton Woods Objectives; Underdeveloped Countries—The Theory and Practice of Technical Assistance; United States Demand for Imports; Interregional Analysis and Regional Development; Governmental Economic Activity; Development of Economic Thought; Technology; Economics in the Curricula of Agricultural Colleges in Canada and the United States; Sociology and Economics; Recent Developments in Mathematical Economics and Econometrics; Research; Long-run Effects of Full Employment on the Labor Market; Theory of Income Distribution; Distribution of Government Burdens and Benefits; Distribution and Utilization of Natural Gas. Pp. 612.

Supplement No. 2—Graduate Education in Economics. Pp. 223.

Supplement No. 3—Handbook. Pp. 187.

Volume XLIV, 1954

The American Economic Review, March, June, September, and December

Supplement (No. 2—May).—Sixty-sixth Annual Meeting:

Papers and Proceedings: Fundamental Characteristics of the American Economy; An Appraisal of Economic Change; Factor Markets Versus Product Markets; Farm Prices and Farm Incomes in American Agriculture; Industrial Pricing; Technological Progress and Economic Institutions; Growth Decisions in the American Economy; Diminishing Inequality in Personal Income Distribution; Wage Determination in the American Economy; Alternative Possibilities of Inflationary Pressures and Higher Cost Bottlenecks in an Economy of Large Bargaining Units and of Less Than Pure and Perfect Competition in the Marketing of Products; Regional Wage Differentials in an Economy of Large Bargaining Units and Less Than Pure and Perfect Competition in the Marketing of Products; Automaticity of Full Employment Under the Assumption of Diminished Defense Expenditures; Institutional Aspects of Savings and Investment; Role of Corporate Taxation in the American Economy; Theory of International Trade in a World of Trade Barriers and Controls and of Variegated National Economic Systems; Economic Doctrines Implied in the Reports of the U.N. and IBRD on Underdeveloped Countries; Corporate International Investment Policies and Programs; Economic Implications of an Aging Population; Report on Graduate Training in Economics; Economics in General Education; Economic and Regulatory Problems in the Broadcasting Industry; National Transportation Policy, Pp. 765.

Supplement.—Supplement to 1953 Handbook. Pp. 11.

Volume XLV, 1955

The American Economic Review, March, June, September, and December
Supplement (No. 2—May).—Sixty-seventh Annual Meeting:

Papers and Proceedings: International Flow of Economic Ideas; Development Policy in Underdeveloped Countries; Economic Development—Case Studies; Regional Economics; Urbanization and Industrialization of the Labor Force in a Developing Economy; Models of Economic Growth; Current Problems in Agricultural Economics; Long-term Trends in International Trade; Economic Research and Public Policy; Economic Stabilization, Forecasting, and the Political Process; Cyclical Experience in the Postwar Period; Debt Management and Monetary Policy; Taxation and Income Distribution; Concepts of Competition and Monopoly; Impact of Antitrust Laws; Research on the Business Firm; Price and Wage Flexibility; Pricing in Transportation and Public Utilities. Pp. 711.

Supplement.—Supplement to 1953 Handbook. Pp. 8.

Volume XLVI, 1956

The American Economic Review, March, June, September, and December
Supplement (No. 2—May).—Sixty-eighth Annual Meeting:

Papers and Proceedings: Statement of the Problem of Keeping the U.S. Economy Moving Forward, But Steadily; Production and Consumption Economics of Economic Growth; Income Distribution Aspects of Expanding Production and Consumption; Increase of Consumption Part of Economic Growth; Government Expenditures and Economic Growth; Monetary Role in Balanced Economic Growth; Shortening Work Week as a Component of Economic Growth; Highway Development and Financing; Unemployment as a Phase of Economic Growth; Urban Growth and Development; Population Specter—Rapidly Declining Death Rates in Densely Populated Countries; Current Economic Thought and Its Application and Methodology in Continental Europe; Economic Thought and Its Application and Methodology in the East; Economic Potentials of Latin America; Economic Potentials of Africa; Report of the Attorney General's Committee on Antitrust Policy; Changing Patterns of Competition in Transportation and Other Utility Lines; Economics in the Curricula of Schools of Business. Pp. 651.

Volume XLVII, 1957

The American Economic Review, March, June, September, and December
Supplement (No. 2—May).—Sixty-ninth Annual Meeting:

Papers and Proceedings: Institutional Economics; Economic Growth and

Development; Keynesian Economics after Twenty Years; Employment Act
in the Economic Thinking of Our Times; Business Fluctuations and In-
flation; Government Taxing and Spending; Developments in Banking and
Monetary Policy; Price and Competitive Aspects of the Distributive Trades;
Monopoly Problem as Seen by Social Scientists; Impact of Some New
Developments in Economic Theory; Economics and Changing Technology;
Power and Public Utility Problems; Transportation Problems; Social
Security and Welfare; Income Distribution; Consumers in the American
Economy; International Economics; Soviet Economic Developments; Eco-
nomics in Schools. Pp. 754.
Supplement (No. 4—July).—Handbook. Pp. 522.†
  † This issue is not included in the price of the volume.

Volume XLVIII, 1958

The American Economic Review, March, June, September, and December
Supplement (No. 2—May).—Seventieth Annual Meeting:
    Papers and Proceedings: Veblen Centenary Round Table; Democracy and
    Trade-Unionism; Agenda for a National Monetary Commission; Is Another
    Major Business Contraction Likely; Monetary Analysis and the Flow of
    Funds; Economic Projections and a Statistical Contribution to Price The-
    ory; Statistical Cost Functions; Trends in Capital Investment and Capac-
    ity; Income and Consumption; State and Local Public Finance; Agricul-
    tural Parity; Further Explorations in Monopolistic-Competitive Price
    Theory; Petroleum and Natural Gas and the Public Interest; Critical Eval-
    uation of Public Regulation of Independent Commissions; Current Eco-
    nomic Questions Relating to Western Europe; Measuring Production in the
    U.S.S.R.; Selected Papers—American Economic Association Competition.
    Pp. 677.
Supplement.—Supplement to 1956 Handbook. Pp. 32.

Volume XLIX, 1959

The American Economic Review, March, June, September, and December
Supplement (No. 2—May).—Seventy-first Annual Meeting:
    Papers and Proceedings: Soviet Economic Trends and Prospects; Soviet
    Economic Planning; Non-Russian Communist Economies; Fundamentals
    of Economic Progress in Underdeveloped Countries; Special Problems
    Facing Underdeveloped Countries; Role and Character of Foreign Aid;
    International Trade and Payments in an Era of Coexistence; Maintaining
    Full Employment and Economic Stability; Balanced Economic Growth
    in History—A Critique; Economics of Government Expenditures; Power
    Blocs and the Operation of Economic Forces; Administered Prices Re-
    considered; Studies in the Classical Economics; Selected Problems in Eco-
    nomic Theory; Open Competition; Organization and Financing of Economic
    Research; Market for Economists. Pp. 689.

Volume L, 1960

The American Economic Review, March, June, September, and December
Supplement (No. 2—May).—Seventy-second Annual Meeting:
    Papers and Proceedings: Standards for the Performance of Our Economic
    System; Relations Between Economic Theory and Economic Policy; Incen-
    tives and Economic Growth: Changing Roles and Public Policies; Prob-
    lems of Achieving and Maintaining: a High Rate of Economic Growth,
    Full Employment, and a Stable Price Level; Problem of Raising Incomes
    in Lagging Sectors of the Economy; Problem of International Harmony:
    Economic Policies for a Lasting Peace; Problem of Social Priorities; In-
    vesting in Education and Research; Facilitating Movements of Labor Out
    of Agriculture; Reducing Impediments to Foreign Trade; Reforming the
    Tax System; Improving the Efficiency of the Transportation and Utilities
    Systems; Research: on Theory of the Firm, on Income, Consumption, and
    Savings, and on Economic Development. Pp. 745.

Volume LI, 1961

The American Economic Review, **March, June, September, and **December; each 1.50

Supplement (No. 2—May).—Seventy-third Annual Meeting:
Papers and Proceedings: Monetary Theory—New and Old Looks; Macroeconomic Theories of Income Distribution; Capital Theory; Managerial Economics—A New Frontier; Frontiers in Uncertainty Theory—The Evidence of Futures Markets; Distribution Costs—Concepts and Measures; Antitrust Problems; Economic Analysis of Urban Problems; Public Utilities and Transportation; Wheat—A Permanent Need for a Farm Program; Problems of Economic Instability in Other Countries; Balance of Payments of the United States—Problems and Prospects; Economics and National Security; Economic Development in Mainland China; Influence of Moral and Social Responsibility on Economic Behavior; Economic Education—Challenge to Our Profession. Pp. 675. $3.00

Volume LII, 1962

The American Economic Review, **March, June, September, and December; each 1.50
**Supplement (No. 2—May).—Seventy-fourth Annual Meeting:
Papers and Proceedings: New Developments in the Theory of the Firm; Problems of Economic Development: Lagging U.S. Growth Rate; International Transmission of Business Cycles—Problems and Policies; Soviet Economic Planning; Economics of Research and Development; Studies in Business Behavior; Economic Behavior of Families; Reappraisal of the Doctrine of Consumer Sovereignty; Report of the Commission on Money and Credit; Tax Problems; Systems of Economic Accounts and Analysis for Urban Regions; Role of Transportation in Economic Development; Transportation Problems in the American Economy; Economics of Water Resource Use; Teaching of Economics; Market for Economists. Pp. 615.   3.00

Volume LIII, 1963
($8.00 a Volume)

The American Economic Review, March, June, September, and December; each 2.00
Supplement (No. 2—May).—Seventy-fifth Annual Meeting:
Papers and Proceedings: Richard T. Ely Lecture; Public Policies with Respect to Private Business; International Commodity Stabilization; Conditions of International Monetary Equilibrium; Problems of Regional Integration; Problems of Methodology; Topics in Economic Theory; Industrial Capacity; Tax Reform; Financial Institutions and Monetary Policy—A Re-examination of Their Interrelationship; Defense and Disarmament; Pricing and Resource Allocation in Transportation and Public Utilities; Postwar Growth in the United States in the Light of the Long-swing Hypothesis; Economic Trends and Prospects in the U.S.S.R. and Eastern Europe; Japanese Economic Development; Economic Development and the Population Problem; Economic Education. Pp. 753.   4.00
Supplement No. 2.—Economics in the Schools. Pp. 27.   .50

Volume LIV, 1964

The American Economic Review, March, June, **September, and December; each   2.00
Supplement (No. 1—January).—Handbook. Pp. 464.†   8.00
Supplement (No. 3—May).—Seventy-sixth Annual Meeting:
Papers and Proceedings: Richard T. Ely Lecture; Theory of Monopolistic Competition after Thirty Years; Principles of Efficiency; Efficiency in Agriculture; Efficiency in the Labor Markets; Regulated Industries; Efficiency in the Use of Natural Resources; Efficiency in the Government Sector; Problems in Taxation; Financial Factors in Business Cycles; Reappraisals in American Economic History; Comparative Costs and Economic Development; Canadian-American Economic Relations—Efficiency and Independence; Efficiency of the Soviet Economy; Invited Dissertations, I and II; Efficiency in the Teaching of Economics. Pp. 698.   4.00
Supplement No. 2—Surveys of Foreign Postwar Developments in Economic Thought. Pp. 55.   1.00
Supplement No. 3—Surveys of Foreign Postwar Developments in Economic Thought. Pp. 101.   1.00
  † This issue is not included in the price of the volume.

Volume LV, 1965

The American Economic Review, **March, June, September, and December; each $2.00
Supplement (No. 2—May).—Seventy-seventh Annual Meeting:

Papers and Proceedings: Richard T. Ely Lecture; Economic Growth—
International Comparisons; Comparative Economic Systems—National-
ized Industry; Economic History—Its Contribution to Economic Edu-
cation, Research, and Policy; Contribution of the History of Economic
Thought to the Understanding of Economic Theory, Economic History,
and the History of Economic Policy; Evolving International Monetary
Mechanism; Domestic Implications of the Evolving International Mone-
tary Mechanism; Survey Research—Three Surveys, Findings and Implica-
tions for Theory and Policy; Recent Capital and Production Theory; Techno-
logical Change—Stimuli, Constraints, Returns; New National Econometric
Model—Its Application; Invited Student Dissertations; Defense Economics
—Applying Economic Criteria; Nonmarket Decision Making; Economic
Theory and Nonprofit Enterprise; Economics of Poverty; Economic Educa-
tion—Experiments in the Teaching of Economics. Pp. 652.   4.00

Supplement (December).—The Structure of Economists' Employment and Salaries,
1964. Pp. 98.   1.00

Volume LVI, 1966
($10.00 a Volume)

The American Economic Review, March, June, September, and December; each 2.00
Supplement (No. 2—May).—Seventy-eighth Annual Meeting:

Papers and Proceedings: Richard T. Ely Lecture; Allocation and Distribu-
tion Theory—Technological Innovation and Progress; Capital Theory—
Technical Progress and Capital Structure; Economic Development—Ad-
vanced Technology for Poor Countries; Knowledge, Information, and In-
novation in the Soviet Economy; Money and Banking—Innovations in
Finance; Public Finance—Promotion of Knowledge Production and Inno-
vation; International Economics—Progress and Transfer of Technical
Knowledge; Antitrust and Patent Laws—Effects on Innovation; Public Reg-
ulation—The Impact of Changing Technology; Economics of Education;
Economics of Publishing; Economics of Broadcasting and Advertising;
Economics of Science Policy; Production and Use of Economic Knowl-
edge; Labor Economics—Effects of More Knowledge. Pp. 670.   4.00

Supplement (September).—Handbook. Pp. 184.   2.50

Volume LVII, 1967

The American Economic Review, March, June, September, and December; each 2.00
Supplement (No. 2—May).—Seventy-ninth Annual Meeting:

Papers and Proceedings: Richard T. Ely Lecture; Military Manpower Pro-
curement; Cost-Benefit Analysis for Government Decisions; Economics of
Health; Economic Analysis of Water Resource Problems; Transportation
and Patterns of Urban Development; Antitrust and Monopoly; Economic
History; Economic Development; Property Rights and Behavior; Topics in
Money; Aggregate Models; Measurement of Price Change; Invited Doctoral
Dissertations; Das Kapital—A Centenary Appreciation; Efficiency of Ed-
ucation in Economics. Pp. 741.   4.00

The titles of papers published in the *Papers and Proceedings,* as well as the titles of
articles in the *American Economic Review,* are indexed in the *Reader's Guide* to 1953
and the *International Index* thereafter.

* * *

Address all inquiries regarding Vols. I-L (1911-60) of the *American Economic Review*
to Walter J. Johnson, Inc., 111 Fifth Avenue, New York, New York, 10003. Inquiries
regarding LI (1961) and subsequent issues of the *Review* and other publications should
be addressed to the American Economic Association office.

* * *

The American Economic Association, founded, among other purposes, for the "en-
couragement of economic research" and "the encouragement of perfect freedom of eco-

nomic discussion," has over fifteen thousand members, including public and professional men and most of the leading students of political economy in America. Membership dues are ten dollars a year. Each member receives all current reports and publications of the Association. In addition to members, there are some six thousand subscribers to the *American Economic Review*, institutional and individual. Applications for membership and other inquiries should be sent to:

<div align="center">

Harold F. Williamson

SECRETARY OF THE AMERICAN ECONOMIC ASSOCIATION

629 Noyes Street, Evanston, Illinois 60201

</div>

\

BARGAINING FOR HEALTH

Labor Unions, Health Insurance, and Medical Care

by Raymond Munts

Examines the evolution of health bargaining detailing the experience in the clothing, coal, auto, and steel industries; explores the implications of "community health bargaining" not only for unions but also for employers, insurance carriers, hospitals, doctors, and medical service plans; and summarizes the contribution of labor to evolving medical economics.

330 pages $7.50
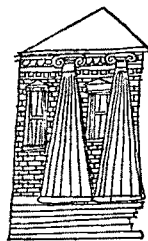
SOVIET METAL-FABRICATING AND ECONOMIC DEVELOPMENT

Practice versus Policy

by David Granick

A study of selected problems of economic development based on the Soviet experience. In order to evaluate both the short- and long-run effects of decisions and organizational traits developed under the first Five Year Plan, it examines in detail the metal-fabricating industries during these key years and traces into the early 1960's the effects of Plan policies.

382 pages $8.50

WILLIAM MORRIS LEISERSON

A Biography

by J. Michael Eisner

An outstanding mediator and arbitrator of labor disputes during the first half of the twentieth century, Leiserson left a valuable legacy in the field of industrial relations—some fifty years of action, policy, precedent, and an abiding faith in the efficacy of collective bargaining.

154 pages, 8 illus. $6.75

THE PORT OF MILWAUKEE

An Economic Review

by Eric Schenker, with the collaboration of John Welcher and John Wilson

Documents the changing character of the port's operations as the volume of its overseas trade increases; investigates the traffic pattern of its hinterland; and predicts a rising role for Milwaukee as a leading port on the nation's fourth seacoast.

about 208 pages, 9 illus. $7.50

At your bookseller or write direct.

THE UNIVERSITY OF WISCONSIN PRESS
BOX 1379   MADISON, WISCONSIN 53701

MODERN ECONOMICS: AN INTRODUCTION

Lorie Tarshis
Stanford University
814 pages   1967 - $8.95

"The book provides an excellent explanation of the market mecha-
nism in operation under a full range of situations other than just per-
fect competition and monopoly. The inclusion of the 'money creation'
process as a product of financial intermediation is also a welcome
addition to improve the student's understanding of the operation of
the monetary sector. It is one of the finest textbooks I have seen."—
John Iacobelli, *The University of Texas.*

GUIDED REVIEW AND WORKBOOK
for Tarshis: MODERN ECONOMICS:
AN INTRODUCTION

Gordon C. Winston and Stephen O'Brien
both of Williams College

475 pages   8½" x 11"   Paper.   April 1967

A valuable adjunct to the Tarshis text, the *Guided Review and Work-
book* contains a Mathematical Preface explaining the uses of mathe-
matics in economics; a Programmed Review for each chapter; and an
abundance of true-false, multiple choice, and essay questions. An *In-
structor's Manual* supplies answers to questions and solutions to prob-
lems.

In preparation for instructors who use the Tarshis text is a *set of ques-
tions for quizzes and tests*—10 to 12 true-false and 5 to 6 multiple
choice—for each chapter in the text.

ECONOMIC ISSUES AND POLICIES
Readings in Introductory Economics—Second Edition

Edited by Arthur L. Grey, Jr., University of Washington
and John E. Elliott, University of Southern California

548 pages   1965   Paper $4.75

"Appropriate stress on policy and balanced selection make it an ideal
supplement for principles courses. High caliber selections and not
unduly polemical. Retains best of first edition."
                              —John R. Eriksson, *Williams College*

of exceptional merit

<table>
<tr><td>

A HISTORY OF ECONOMIC THOUGHT

|N FRED BELL
versity of Illinois

Published—Second Edition

book traces the development
onomic thought in Western
ization. Concentrating
conomists who have made
ficant contributions and on
ols of thought with lasting
ence, it helps the student
mprehend and appraise
mporary economic thinking.
ological, it clarifies the
ionary process in economic
/ and practice, and points
e interrelationships between
eas of thinkers in
nt periods. The book
zes and evaluates the
ble doctrines, discusses
nfluence on national
s, and relates them to
t-day problems. This Second
n includes new interpretive
ial throughout; a broader
ssion of the influence of
alist criticism on subsequent
ought; contemporary American
contributions; and a chapter on
the evolution of mathematical
economics.

2nd Ed., 1967. 745 pp. $9.00

</td><td>

ECONOMIC AND SOCIAL SECURITY

JOHN G. TURNBULL and
C. ARTHUR WILLIAMS, JR.
—both University of Minnesota;
and **EARL F. CHEIT**, University of
California, Berkeley

Just Published—Third Edition

Designed for students of
social insurance and social
security, this book offers
thorough discussion of various
public and private measures for
extending economic security.
Following the successful pattern
established in previous editions,
Third Edition retains the use of
analytical tools in evaluating
private and social insurances
and assistances involved in
such problems as old age and
death, occupational disability
and non-occupational illness,
unemployment, and substandard
working conditions. New stress
is placed on the problem of
poverty. Emphasis throughout is
on principles and understanding,
and any abstract discussion of
statistical detail or
administrative complexity is
avoided.

3rd Ed., 1967. 671 pp. $9.00

</td></tr>
</table>

······15 East 26th Street / New York, N.Y. 10010